NON-STATIONARY TRANSPORT EFFECTS IN DEEP SUB-MICRON CHANNEL Si MOSFET'S

by

Ghavam Ghavami Shahidi

B.S. Massachusetts Institute of Technology (1981)
M.S. California Institute of Technology (1982)
M.S. Massachusetts Institute of Technology (1984)
E.E. Massachusetts Institute of Technology (1984)

Submitted in Partial Fulfillment of

of the Requirements for the Degree of

Doctor of Philosophy

in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

January 1989

©Massachusetts Institute of Technology, 1989

Signature of Author ______________________________
Department of Electrical Engineering and Computer Science
January 1989

Certified by ______________________________
Professor Dimitri A. Antoniadis
Thesis Supervisor

Certified by ______________________________
Professor Henry I. Smith
Thesis Supervisor

Accepted by ______________________________
Professor Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AUG 07 1991

LIBRARIES

NON-STATIONARY TRANSPORT EFFECTS IN DEEP SUB-MICRON CHANNEL Si MOSFET'S

by

Ghavam Ghavami Shahidi

Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in Electrical Engineering and Computer Science

Abstract

This thesis is concerned with conduction in high electric field and high field gradient regions; i.e. modifications in transport due to presence of sharp changes in drift field. As device dimensions are reduced into the submicron range, such devices operate under high electric field and field gradient. Questions about the validity of basic drift-diffusion equation have been raised. The most fundamental issue has been the functional dependence of transport coefficients; i.e. if the transport coefficients (mobility, diffusitivity) are local function of electric field (as is usually incorporated into drift-diffusion) or they are functions of some other variable, which in turn was modified by the field (i.e. energy or temperature). The origin of these issues has been the theoretical prediction of velocity overshoot (by means of Monte Carlo simulation). These effects are caused by the lack of steady state equilibrium between the carrier temperature and the local field when the field gradient is sufficiently large, i.e. the carriers are in a "non-stationary" equilibrium with the applied field. This thesis is concerned with such effects, and their importance.

The ideal device to study such effects is the deep-submicron channel MOSFET: It has a high field region in the channel and a source of cold carriers in the source. Carrier are injected from the source into the high field channel. If the device is sufficiently short, the electric fields in the channel are sufficiently high to observe some of these non-stationary effects. In order to observe these effects at higher temperatures, carrier scattering in the channel must be reduced; i.e. the channel mobility must be high. The first part of this thesis deals with the fabrication of high mobility deep-submicron channel MOSFETs. The basic process already existed, and some slight modifications were used to obtain devices with higher mobility and smaller parasitic resistances.

Once the high mobility MOSFETs were fabricated, velocity overshoot were readily observed at room temperature and 77 K for the first time. This observation proves the existence of non-stationary transport. The magnitude of such an effect is small at room temperature (20% over the bulk value) and high at 77 K (80% over the bulk value). It is not clear if a complicated alteration of simple drift-diffusion is necessary for reliable device modeling. Velocity overshoot is caused by the reduced carrier temperature in the channel

(relative to the one corresponding to local field). Therefore, this prompted investigation of hot electron effects in deep submicron range. Specifically, hot-electron generated substrate current was measured, and a reduction in the substrate current was observed for devices with channel length L< 0.15 μm. An increased amount of substrate current reduction was observed at 77 K in short channel devices. It is proposed that a reduction of carrier temperature due to rapid variation in the lateral electric field and a reduction in the carrier density in the high field region because of velocity overshoot, is responsible for the reduction in the substrate current generation in the deep submicron channel length range.

To reduce short channel effects in deep submicron channel devices, indium was used as channel implant for the first time. It is possible to keep the surface doping low and bring the heavy p^+ doping very close to the surface, thus reducing the short channel effects. At the same time, in these devices, both the velocity overshoot and the reduction in substrate current was observed.

These observations, prove the exictance of the so called non-stationary effects, and highlight the need for their inclusion (through higher moments of Boltzmann Equation or use of Monte Carlo method) in modeling of short channel devices.

Thesis Supervisor: Dr. Dimitri A. Antoniadis
Title: Professor of Electrical Engineering

Thesis Supervisor: Dr. Henry I. Smith
Title: Professor of Electrical Engineering

Acknowledgment

First I would like to thank my thesis supervisors, Professors Dimitri A. Antoniadis and Henry I. Smith for their invaluable guidance, support, and patiance throught this work. Professor Antoniadis had great insight, and always directed me toward the right direction. Professor Smith's guidance in fabrication and excitment about the work were indispensible. I am indebted to them for ever.

I would like to thank the readers, Professors Marc A. Kastner and Terry P. Orlando for their suggestions. This work could not have been done without the smooth operations of Submicron Strautures Lab and Microsystems Lab. I am grateful to Jimmy Carter for his valuable help over many years of both this work and M.S. work. Joe Walsh, Mike Schroth, Paul Maciel , and Bayla Fine always were more than willing (and able) to help. I thank Tim McClure, Robert Cuikay and Velma McClure for their expert assistance throught different stages of fabrication. Phyllis deFano and Cristina Gordy were always more than helpful with the bureaucratic matters.

I am grateful to Jacob Jarvis for setting up MINIMOS and *always* going out of his way to help, Merit Huang for his excellent plotting routine, and Steve Chou with for his initial help with the masks.

I thank my friends and fellow graduate students: Mehran Mehregany, Anthony Yen, Philip Bagwell, Khalid Ismail, Duane Boning, Alberto Moel, Yau-Chin Que, William Chu, and Chun The, for making our groups such a pleaseant place to work.

This work was supported by the Joint Services Electronics Program, and Air Force Office of Sponsered Research, contract AFOSR-85-0376.

To:

Dr. Amir M. Farhadi

Dr. Mary Farhadi

Contents

Abstract 2

Acknowledgement 4

List of Figures 7

List of Tables 11

1. Introduction 12

2. MOSFET Device Physics 16

2.1.1 MOSFET Structure 16

2.1.2 MOS Capacitor Structure 18

2.1.3 MOSFET Operation 23

2.2 Mobility in Inversion Layers 25

2.3 Scaling 28

2.4.1 Short Channel Effects 31

2.4.2 Charge Sharing Model 31

2.5.1 Velocity Saturation 35

2.5.2 MOSFET with velocity Saturation 36

3. Fabrication 38

3.1 Introduction 38

3.2.1 Fabrication of X-ray Mask 39

3.2.2 X-ray Exposure 43

3.3 Device Fabrication 45

3.4 Results 48

3.4.1 Contact and Sheet Resistivities 49

3.4.2 Gate Oxide 49

3.4.3 Process Result Summary 49

4. Non-Stationary Transport 52

4.1 Non-Stationary Transport 53
4.2 The Boltzmann Equation 54
4.3 The Hydridynamic Equations 56
4.3.1 Complete Hydrodynamic Equations 57
4.3.2 Solution of Coupled Equations 61

5. Velocity Overshoot 70
5.1 Velocity Overshoot 71
5.2.1 Experimental Results 73
5.2.2 Observation of Overshoot 77

6. Hot Electron Effects 82
6.1 Introduction 82
6.2 Avalanch Multipication 83
6.3.1 Lateral Electric Field 84
6.3.3 Derivation of Lateral Field 85
6.4 Substrate Current 89
6.5 Experimental Results at 300 K 91
6.6 Experimental Results at 77 K 98

7. Indium Implanted MOSFET Channel 102
7.1 Introduction 102
7.2.1 Experimental Results 105
7.2.2 Necessity of Deep Implant 105
7.2.3 Mobility 105
7.2.4 Device Characteristics 108
7.2.5 Hot Electron Behavior 112

8 . Conclusion 116

Appendix A: Laterl Field in MOSFETs 119

Appendix B: Detailed Fabrication Steps 125

Bibligraphy 132

List of Figures

2.1 MOSFET cross section 17

2.2 (a) MOS capacitor cross section. (b) MOS capacitor band diagram 19

2.3 Total charge induced as a function of ψ_S 22

2.4 Cross section of a short channel MOSFET, showing the the charge shared, ΔQ_B 33

3.1 An x-ray mask: A thin polyimide memberane stretched on a vesple ring 41

3.2 SEM micrograph of a (110) Si wafer in which wells were etched. 42

3.3 Steps involved in fabrication of x-ray mask 44

3.4 A 0.12 μm wide line obtained by x-ray exposure. 46

3.5 Movement of dies (relative to the center of x-ray mask), as measured by the stepper. 47

3.6 A 0.15 μm device produced by our process. This device has the highest g_m reported to date, despite its relatively high R_{SD}. 51

4.1 Carrier velocity in region of high carrier gradient as calculated by drift-diffusion and Hyrodynamic equations [Baccarani and Wordeman]. 63

4.2Elecctric field profile used to study non-stationary effect caused by rapid variations in electric field 67

4.3 (a) Carrier velocity in a region of sharp electric field discontinuity (Solid line is exact

and dashed line is using effective D) (b) Carrier temperature in a region of sharp electric field discontinuity 68

4.4 (a) Carrier velocity in a region of sharp electric field discontinuity for different mobilities. (b) Carrier temperature in a region of sharp electric field discontinuity field discontinuity for different mobilities 69

5.1 (a) Lateral field for a 0.5 μm device for V_{DS} = 2.0 V. (b) Carrier velocity for a 0.5 μm device for V_{DS} = 2.0 V 74

5.2 (a) Lateral fields for a 0.1 μm and 0.15 μm device for V_{DS} = 2.0 V. (b) Carrier velocities for a 0.1 μm and 0.15 μm device for V_{DS} = 2.0 V 75

5.3 (a) Channel doping used in observation of velocity overshoot. (b) Threshold voltage vs. channel length for uniform and nonuniform substrate doping, as simulated by a two dimensional device simulator. The surface doping for the nonuniform case is 2×10^{16} cm^{-3}. Experimental results for the nonuniform doping case are also shown. 78

5.5 Charactristics of a 90-nm-channel length device at (a)300 K and (b) 77 K. . . . 79

5.6 (a) Measured transconductances of devices vs channel lengths at 300 K and 77 K. (b) Calculated average velocities at 300 K and 77 K. 81

6.1 Drain region of a MOSFET as modeled by Ko, showing effect of x_j. 88

6.2 (a) Drain and substrate current at room temperature for two long channel MOSFETs. (b) Drain and substrate current at room temperature for two two short channel MOSFETs under identical bias as in (a) 82

6.3 Plots of $log(I_{SUB}/I_D)$ vs. $(V_{DS} - V_{DSAT})^{-1}$ at room temperature for devices with channel lengths from 0.06 μm to 5 μm. The arrows trace the "evolution" of the plots as channel length is decreased. 93

6.4 Plots of E_m, the maximum field in the cahannel vs. $(V_{DS} - V_{DSAT})^{-1}$ at room temperature for devices with channel lengths from 0.1 μm to 5 μm (As calculated by MINIMOS). 95

6.5 The maximum carrier temperauture in the channel at $V_{DS} = 2.0$ V and $V_{GS} = 1.5$ V, at room temperature for devices with channel lengths from 0.08 μm to 5 μm with and without velocity overshoot. 97

6.6 (a) I_D and I_{SUB} vs. V_{DS} at $(V_{GS} - V_T) = 0.5V$ for a long-channel device, L= $2.5\mu m$ ($V_T = 0.19\ V$, at 300 K and $V_T = 0.46\ V$, at 77 K), at room temperature (solid line) and at 77 K (dashed line). (b) I_D and I_{SUB} vs. V_{DS} at $(V_{GS} - V_T) = 0.5V$ for a short-channel device, L= $0.14\mu m$ ($V_T = 0.07\ V$, at 300 K and $V_T = 0.33\ V$, at 77 K). 99

6.7 Plots of $log(I_{SUB}/I_D)$ vs. $(V_{DS} - V_{DSAT})^{-1}$ for $0.14\mu m < L < 5\mu m$, at room temperature and 77 K. Note the enhanced reduction of the substrate current generation at low temperature for short-channel devices. 101

7.1 Concentration of Indium implanted at three different energies and boron at 50 KeV, as simulated by SUPREM. 104

7.2 Reduction in V_T as function of V_{DS} for different values of implant energy. . . 106

7.3 (a): Subthreshold conduction for a submicron MOSFET without the deep implant. (b): Subthreshold conduction for a submicron MOSFET with the deep implant. . 107

7.4 Mobility in a boron doped channel and an Indium doped channel (with twice the dose and much shallower). 109

7.5 (a) Comparision of Inium and boron doped samples (Table 7.1) in terms of body effect, (b) Effect of dose on body effect, and (c) Effect of implanr energy on body effect ($t_{ox} = 5.3nm$) 110

7.6 Plots of $log(I_{SUB}/I_D)$ vs. channel length at $(V_{DS} - V_{DSAT})^{-1} = 0.55\ V^{-1}$ for different values of implant energies. Note the strong dependence of substrate current generation on implant energy for submicron devices. (a) is without the deep boron implant, and (b) is with the deep implant. 113

7.7 The drain and the corresponding substrate current for two short channel devices (L=0.1 and 0.15 μm), at $V_{GS} = 1.2$ V. Note the reduction in the substrate current, despite the increase in I_D. 115

List of Tables

2.1 General Scaling Laws 29

3.1 Improvements Obtained in Process 48

7.1 Comparision of In and B doped Devices 110

List of Figures

2.1 MOSFET cross section 17

2.2 (a) MOS capacitor cross section. (b) MOS capacitor band diagram 19

2.3 Total charge induced as a function of ψ_S 22

2.4 Cross section of a short channel MOSFET, showing the the charge shared, ΔQ_B 33

3.1 An x-ray mask: A thin polyimide memberane stretched on a vesple ring 41

3.2 SEM micrograph of a (110) Si wafer in which wells were etched. 42

3.3 Steps involved in fabrication of x-ray mask 44

3.4 A 0.12 μm wide line obtained by x-ray exposure. 46

3.5 Movement of dies (relative to the center of x-ray mask), as measured by the stepper. 47

3.6 A 0.15 μm device produced by our process. This device has the highest g_m reported to date, despite its relatively high R_{SD}. 51

4.1 Carrier velocity in region of high carrier gradient as calculated by drift-diffusion and Hyrodynamic equations [Baccarani and Wordeman]. 63

4.2 Elecctric field profile used to study non-stationary effect caused by rapid variations in electric field 67

4.3 (a) Carrier velocity in a region of sharp electric field discontinuity (Solid line is exact and dashed line is using effective D) (b) Carrier temperature in a region of sharp electric field discontinuity 68

4.4 (a) Carrier velocity in a region of sharp electric field discontinuity for different mobil-

ities. (b) Carrier temperature in a region of sharp electric field discontinuity field discontinuity for different mobilities 69

5.1 (a) Lateral field for a 0.5 μm device for $V_{DS} = 2.0$ V. (b) Carrier velocity for a 0.5 μm device for $V_{DS} = 2.0$ V 74

5.2 (a) Lateral fields for a 0.1 μm and 0.15 μm device for $V_{DS} = 2.0$ V. (b) Carrier velocities for a 0.1 μm and 0.15 μm device for $V_{DS} = 2.0$ V 75

5.3 (a) Channel doping used in observation of velocity overshoot. (b) Threshold voltage vs. channel length for uniform and nonuniform substrate doping, as simulated by a two dimensional device simulator. The surface doping for the nonuniform case is 2×10^{16} cm^{-3}. Experimental results for the nonuniform doping case are also shown. 78

5.5 Charactristics of a 90-nm-channel length device at (a)300 K and (b) 77 K. . . . 79

5.6 (a) Measured transconductances of devices vs channel lengths at 300 K and 77 K. (b) Calculated average velocities at 300 K and 77 K. 81

6.1 Drain region of a MOSFET as modeled by Ko, showing effect of x_j. 88

6.2 (a) Drain and substrate current at room temperature for two long channel MOSFETs. (b) Drain and substrate current at room temperature for two two short channel MOSFETs under identical bias as in (a) 82

6.3 Plots of $log(I_{SUB}/I_D)$ vs. $(V_{DS} - V_{DSAT})^{-1}$ at room temperature for devices with channel lengths from 0.06 μm to 5 μm. The arrows trace the "evolution" of the plots as channel length is decreased. 93

6.4 Plots of E_m, the maximum field in the cahannel vs. $(V_{DS} - V_{DSAT})^{-1}$ at room temperature for devices with channel lengths from 0.1 μm to 5 μm (As calculated by MINIMOS). 95

6.5 The maximum carrier temperauture in the channel at $V_{DS} = 2.0$ V and $V_{GS} = 1.5$ V, at room temperature for devices with channel lengths from 0.08 μm to 5 μm with and without velocity overshoot. 97

6.6 (a) I_D and I_{SUB} vs. V_{DS} at $(V_{GS} - V_T) = 0.5V$ for a long channel device, L= 2.5μm ($V_T = 0.19$ V, at 300 K and $V_T = 0.46$ V, at 77 K), at room temperature (solid line) and

at 77 K (dashed line). (b) I_D and I_{SUB} vs. V_{DS} at $(V_{GS} - V_T) = 0.5V$ for a short-channel device, L= $0.14\mu m$ ($V_T = 0.07$ V, at 300 K and $V_T = 0.33$ V, at 77 K). 99

6.7 Plots of $log(I_{SUB}/I_D)$ vs. $(V_{DS} - V_{DSAT})^{-1}$ for $0.14\mu m < L < 5\mu m$, at room temperature and 77 K. Note the enhanced reduction of the substrate current generation at low temperature for short-channel devices. 101

7.1 Concentration of Indium implanted at three different energies and boron at 50 KeV, as simulated by SUPREM. 104

7.2 Reduction in V_T as function of V_{DS} for different values of implant energy. . . 106

7.3 (a): Subthreshold conduction for a submicron MOSFET without the deep implant. (b): Subthreshold conduction for a submicron MOSFET with the deep implant. . 107

7.4 Mobility in a boron doped channel and an Indium doped channel (with twice the dose and much shallower). 109

7.5 (a) Comparision of Inium and boron doped samples (Table 7.1) in terms of body effect, (b) Effect of dose on body effect, and (c) Effect of implanr energy on body effect ($t_{ox} = 5.3nm$) 110

7.6 Plots of $log(I_{SUB}/I_D)$ vs. channel length at $(V_{DS} - V_{DSAT})^{-1} = 0.55$ V^{-1} for different values of implant energies. Note the strong dependence of substrate current generation on implant energy for submicron devices. (a) is without the deep boron implant, and (b) is with the deep implant. 113

7.7 The drain and the corresponding substrate current for two short channel devices (L=0.1 and 0.15 μm), at $V_{GS} = 1.2$ V. Note the reduction in the substrate current, despite the increase in I_D. 115

List of Tables

2.1 General Scaling Laws 29

3.1 Improvements Obtained in Process 48

7.1 Comparision of In and B doped Devices 110

Chapter 1

Introduction

Probably the most widely used electronic component is the silicon metal-oxide field effect transistor (MOSFET): It is the basic building block of logic and memory circuits in most computers and digital systems.

It is possible to form two dimensional (and one dimensional) electronic systems using the inversion layer of a MOSFET, and to control the carrier density in the 2-dimensional system over many orders of magnitude. As a result, the Si MOSFET has become one of the basic tools used in the study of electronic transport effects in so-called 2-dimensional electron gas systems, prticulary at cryogenic temperatures.

Since the invention of the integrated circuit, there has been an steady effort to increase the speed, density, and number of devices per chip. As technology has advanced, the smallest dimension in a device has been reduced exponentially with time. As the smallest device dimensions are reduced below 1 μm, straight forward device scaling seems questionable: The major concerns are device behavior as dimensions are reduced, and yield and manufacturability of the system containing small devices. A few years ago, 1 μm was touted to be the limit for the smallest practical device di-

mension. Now use of 0.1 μm channel length devices not only seems to be possible, but inevitable in the near future.

Development of well-behaved short channel MOSFETs is essensial for scaling. Smaller devices are faster, and systems containing them are denser and faster. For practical applications, small MOSFETs should have a clear saturation region of operation (low output conductance), and it must be possible to turn them off sharply below threshold (high subthreshold slope). To achieve this, accurate modeling , device design, and processing are necessary.

As the device dimensions are reduced into the submicron range, in addition to traditionaly studied short channel effects (effects caused by the drain and source junction control of the channel charge in the MOSFET), a number of new physical phenomena take place. Under these conditions, some of the assumptions used in modeling the longer channel devices break down. The origin of this discrepency is that in small devices, the carrier relaxation lengths and the distances over which the electric field and carrier density change, become comparable to the device length. In other words, the distances over which the carriers feel the strong electric field, and adjust to it, are close to the device length, i.e. the carriers are not in stationary balance with the electric field over a significant part of the channel. The effects caused by lack of equilibrium between the high electric field and the carrier properties are called non-stationary effects.

This thesis is a study of such non-stationary effects. The basic tool used in this work is the deep-submicron channel length Si MOSFETs. The short channel MOS-

FET is an ideal device for the study of such phenomena. It has a region of very high electric field (channel) and an injector of cold carriers (source). Many transport-related variables can be monitored over a wide lattice and carrier temperature, electric field strength, carrier density, and, most importantly, length scales. To maximize the amount of the non-stationary effects, one has to increase the relaxation lengths. Therefore, carrier scattering in the channel of the MOSFET has been minimized. Among the non-stationary effects, probably the most prominent is velocity overshoot. This phenomenon has been oberved previously at 4.2 K [Chou, 1985]. The major goal here was to observe it phenomenon at higher temperatures. Therefore, short channel MOSFETs with high mobility had to be fabricated as the first major part of the work. Once high mobility MOSFETs were obtained, velocity overshoot at room temperature was readily observed. A more interesting (and probably more important) effect observed, is reduction in hot electron effects. As there is less interaction between the carriers and the lattice (because of the reduced carrier temperature) such a reduction was expected. As the channel length was reduced into the deep-submicron range, a reduction in the impact ionization-induced substrate current was observed. It seems that velocity overshoot and reduction in hot electron effects are two closely related phenomena. Both these effects not only prove the existence of non-stationary transport, which occurs over short times and/or distances, and highlight the need for inclusion of a more complete transport model in device simulators, used for modeling short channel devices.

As it turns out, many of the non-stationary effects discussed above manifest them-

selves in the device performance, in devices with L$\approx$ 0.1 μm. Therefore, it is important to observe these effects in devices which are well behaved. Thus, throughout this work, the effort has been made to minimize parasitic short channel effects.

Besides this introduction and conclusion, this thesis contains six main chapters. Chapter two is a brief review of MOS device physics. The goal has been to emphasize the limitations and problems of commonly used device models. Chapter three is on device fabrication. The process used in this thesis was already developed. Modifications were made to obtain major improvements in device behavior and yield. Chapter four is a review of non-stationary transport in Si. The emphasis is on a hydrodynamic carrier transport model. The goal has been to gain physical insight into the mechanisms reponsible for non-equilibrium transport. Chapter five presents the results on observation of velocity overshoot in Si at room temperature and 77 K. Chapter six is on hot electron effects and the reduction of such effects in deep submicron devices. Chapter seven is on the use of indium as an alternative species to achieve high mobility and threshold. With indium it is possible to observe all the intersting non-stationary effects in short channel MOSFETs, without the presence of unwanted short channel effects.

Chapter 2

MOSFET Device Physics

This chapter discusses the fundamentals of MOSFET operation. The basic MOS capacitor, the gradual channel approximation and operation of a MOSFET are presented. Next, issues relating to scaling, mobility, short channel effects, and velocity saturation are presented. Emphasis is placed on limitations of the simple MOSFET theory.

2.1 MOSFET Structure

The basic MOS transistor is shown in Fig. 1. The source and drain are heavily doped regions, and the channel is doped with opposite polarity. A gate controls the carrier density on the surface. For the appropriate gate bias, the surface becomes inverted, i.e. it is populated by the same kind of carriers as majority carriers in the source and drain. Thus a conductive path between source and drain is formed, and its conductivity can be modulated by the gate.

Source and drain extend into the z direction, normal to the plane of Fig.2.1. A two dimensional electron gas is formed in the $y - z$ plane. When modeling the MOSFET device operation, the variations along the z-axis are ignored. To be accurate, the full two dimensional Poission equation must be solved (in $x - y$ plane):

$$\nabla^2\psi = \frac{d^2\psi}{d^2x} + \frac{d^2\psi}{d^2y} = -\frac{q}{\epsilon_{Si}}(N - n + p) \qquad (2.1.1)$$

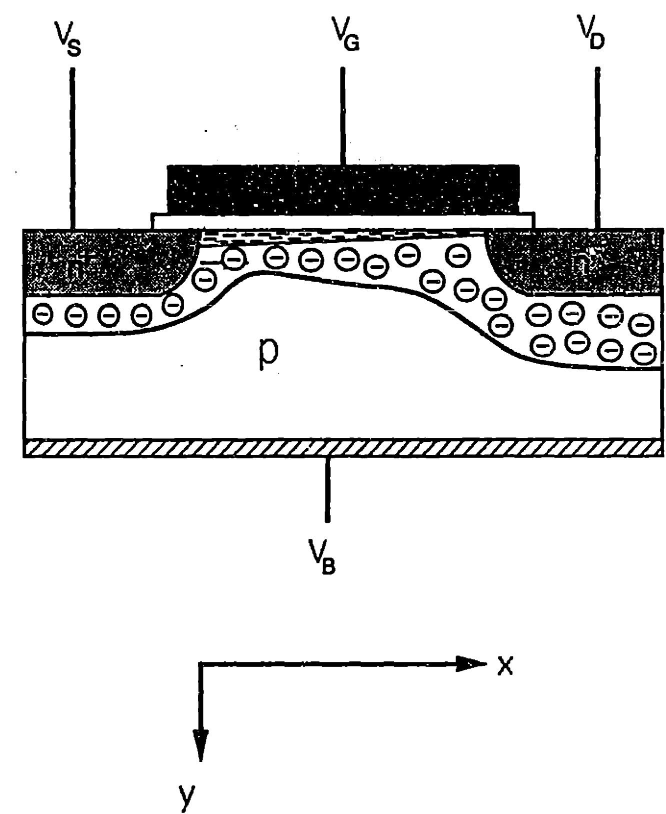


Figure 2.1: MOSFET cross section

where ψ is the electrostatic potential. Solving the full two dimensional equation implies that the charges in the channel contribute to both lateral and normal electric field. These changes are controlled both laterally (through source and drain bias) and vertically (through gate induced electric field). In long channel devices, it can be shown that most of the charge in the channel is controlled by the gate.

$$\frac{d^2\psi}{d^2x} >> \frac{d^2\psi}{d^2y} \qquad (2.1.2)$$

Under this approximation:

$$\frac{d^2\psi}{d^2x} = \frac{-q}{\epsilon_{Si}}(N - n + p) \qquad (2.1.3)$$

is solved, and variations along y are ignored. It is assumed that the gate induced vertical electric field is much greater than the lateral field and the latter is ignored when calculating the charge in the channel. Lateral variation in the electric field is then accounted for by a simple resistive voltage drop along the device and the flow of the I_{SD}. This formulation is widely used and is called the gradual channel approximation (GCA). Within the context of GCA, the lateral field does not induce any charges in the channel. Therefore, two dimensional analytical solutions to the Poission's equation are not required.

2.1.2 MOS Capacitor Structure

Within GCA, first the charge density induced in the inversion layer through the gate bias is obtained. This is the same as the one dimensional MOS capacitor. Fig. 2.2 shows the cross section of the MOS capacitor and the corresponding band diagram. Equation (2.1.3) must be solved.

Introducing the Fermi potential in the semiconductor bulk, ψ_F (and the thermal voltage $\phi_t = \frac{kT}{q}$) [Sze]:

$$N_D^+ - N_A^- = n - p = N_D - N_A = 2n_i \sinh(\frac{\psi_F}{\phi_t}) \qquad (2.1.4)$$

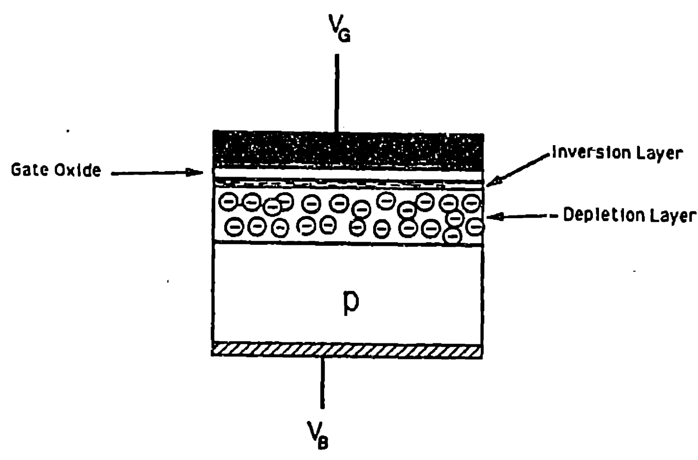


Figure 2.2a: MOS capacitor cross section

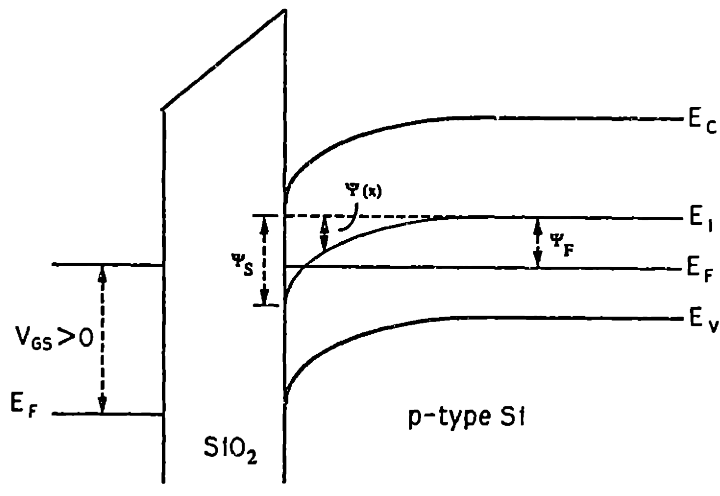


Figure 2.2b: MOS capacitor band diagram

then one obtaines (in the presence of the electrostatic potential ($\psi = \psi(x)$):

$$n = n_i e^{-\frac{(\psi_F - \psi)}{\phi_t}} \tag{2.1.5a}$$

$$p = p_i e^{\frac{(\psi_F - \psi)}{\phi_t}} \tag{2.1.5b}$$

The differential equation for the carrier density in the MOS capacitor becomes:

$$\frac{d^2\psi}{d^2x} = \frac{2qn_i}{\epsilon_{Si}}(\sinh(-\frac{\psi_F}{\phi_t}) - \sinh(\frac{(\psi_F - \psi)}{\phi_t})) \tag{2.1.6}$$

This differntial equation is solved subject to the following boundary conditions:

$$\psi(x = 0) = \psi_S \tag{2.1.7a}$$

$$\psi(x = Large) = 0 \tag{2.1.7b}$$

The first states that the surface potential is determined by the gate bias (the relation will be determined later), and the second boundary condition is just a statement of the charge neutrallity in the bulk.

Integrating 2.1.6 from $x = Large$ to x away from the interface, the electric field becomes:

$$E(\psi(x)) = -\frac{d\psi}{dx} = \pm(\frac{2qn_i}{\epsilon_{Si}})^{1/2}[(N_A - N_D)\psi + n_o\phi_t(e^{-\frac{\psi}{\phi_t}} - 1) + p_o\phi_t(e^{\frac{\psi}{\phi_t}} - 1)]^{1/2} \tag{2.1.8}$$

where n_o and p_0 are the equilibrium values of electron and hole density in the silicon. Knowing the electric field as a function of the potential in the semiconductor, the total induced charge per unit area, Q_C, can be found using:

$$Q_C = \epsilon_{Si}E_{Surface} = \epsilon_{Si}E(\psi = \psi_{Surface}) \tag{2.1.9}$$

The result is then:

$$Q_C = \pm(2qn_i\epsilon_{Si})^{1/2}[(N_A - N_D)\psi_S + n_o\phi_t(e^{-\frac{\psi_S}{\phi_t}} - 1) + p_o\phi_t(e^{\frac{\psi_S}{\phi_t}} - 1)]^{1/2} \tag{2.1.10}$$

where the sign of Q_C agrees with the sign of ψ_S. In Fig. 3, the total charge induced in the semiconductor as a function of ψ_S is ploted.

Equation (2.1.10) describes all the regions of operation of a MOS structure: For p-type silicon, if ψ_S is negative, the second term in the bracket dominates, and the majority holes are present (accumulation). If ψ_s is small and negative, a depletion layer is formed. For large negative ψ_S, the minority electron density dominates, and inversion occurs. All that remaines, is to relate the gate bias to ψ_S. The field in the oxide is:

$$E_{ox} = \frac{\epsilon_{Si}}{\epsilon_{ox}} E_S \tag{2.1.11}$$

The voltage drop in the oxide is:

$$\Delta V_{ox} = \frac{\epsilon_{Si} t_{ox}}{\epsilon_{ox}} E_S \tag{2.1.12}$$

Then:

$$V_G = \psi_s + \Delta V_{ox} \tag{2.1.13}$$

This eqation can be solved for ψ_S for a given V_G, and the total charge in the semiconductor can be found.

The above equation can be solved exatly. It can be seen that, in inversion, the total charge has two parts: First is that of the depletion region and second is that of a very narrow sheet of high density on the surface. This has prompted the introduction of delta-depletion approximation. That is inversion sheet is modeled by an infinitely thin charge sheet. In this model strong inversion occures for $\psi_S = 2\psi_F + \Delta\psi = \phi_B$, where $\Delta\psi$ is of the order of 3-6 ϕ_T [Tsividis]. The extra $\Delta\psi$ is due to the fact that the surface, in strong inversion, gets pinned to ϕ_B and not $2\psi_F$ as many authors have claimed. Total inversion charge density is: $Q_I = C_{ox}(V_{GS} - V_T)$ where the threshold voltage is:

$$V_T = V_{FB} + \phi_B + \gamma(\sqrt{\phi_B} - \sqrt{\phi_B + V_{BS}}) \tag{2.1.14}$$

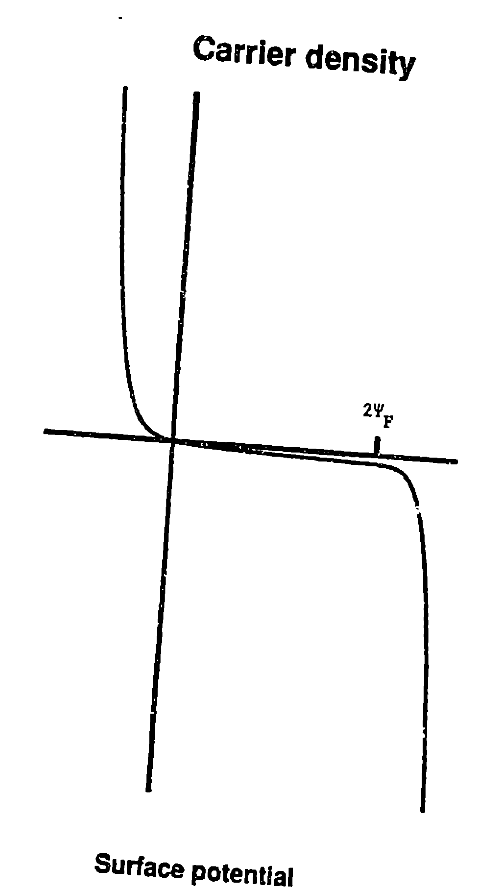


Figure 2.3: Total charge induced as a function of ψ_S

where $\gamma = \sqrt{\frac{2\epsilon_{Si} q N_A}{C_{ox}}}$ is the body effect, $V_{FB} = \phi_{MS} + \phi_B + \frac{Q_T}{C_{ox}}$ is the flat band voltage, ϕ_{MS} is the work function, Q_T is the total charges on the semiconductor surface and in the oxide (mobile and fixed), and V_{BS} is the substrate bias (with respect to the inversion layer). The depletion charge, when in inversion, is given by $Q_{depl} = -qN_A W_{depl}$, where the depletion region width is:

$$W_{depl} = \sqrt{\frac{2\epsilon_{Si}}{qN_A}\phi_B} \tag{2.1.15}$$

2.1.2 MOSFET Operation

In this section, a simplified theory of the p-channel MOS transistor is presented. The effects due to small geometry and variations in the mobility are ignored. The approximations used in GCA are emphasized.

The total current in MOS transistor is made up of drift and diffusion parts:

$$dI_D = dI_{drift} + dI_{diff} \tag{2.1.16}$$

For the drift current at a distance dx below the surface, at location (x, y):

$$dI_{drift} = q\mu n(x,y)dx\frac{\partial\psi(x,y)}{\partial y} \tag{2.1.17a}$$

and the diffusion component is:

$$dI_{diff} = -wDdx\frac{\partial n(x,y)}{\partial y} \tag{2.1.17b}$$

Now since $n(x,y) \propto e^{\psi(x,y)}$, it can be shown that [Tsividis]:

$$I_D = -\frac{W}{L}\int_{V_{SB}}^{V_{DB}} \mu Q_I dV \tag{2.1.18}$$

Where Q_I is the mobile channel charge. This is an exact expression and can be used in all regions of operation. If the lateral field is much smaller than the normal field (GCA), then

the free charge obtained from calculations of one dimensional MOS structure can be used. One can use either the charge obtained through the charge sheet model (i.e. Q_C in (2.1.10)) or use the more exact charge obtained from the exact solotion of Poission's equation in one dimension (Pao-Sah Model). Here, the charge sheet model will be presented.

In charge sheet model the mobile charge density $Q_I(y)$ is:

$$Q_I(y) = -C_{OX}(V_{GB} - V_{FB} - \phi_B - V_{CB} + \frac{Q'_B}{C_{OX}}) \tag{2.1.19}$$

where Q'_B is:

$$Q'_B = -\gamma C_{ox}\sqrt{\phi_B + V_{CB}} \tag{2.1.20}$$

and γ is the body effect, given before. Carrying out the above integration, for the drain current in the linear region [Tsividis]:

$$I_{DS} = \frac{W}{L}\mu C_{ox}[(V_{GS} - V_{FB} - \phi_B)V_{DS} - \frac{V^2{}_{DS}}{2} - \frac{2}{3}\gamma((\phi_B + V_{SB} + V_{DS})^{3/2} - (\phi_B + V_{SB})^{3/2})] \tag{2.1.21}$$

This expression for current can be simplified to:

$$I_{DS} = \frac{W}{L}\mu C_{ox}[(V_{GB} - VT)V_{DS} - 0.5(1+\delta)V^2{}_{DS}] \tag{2.1.22}$$

where δ is related to the *gamma*. The phenemenon of current saturation and pinch-off in the channel are predicted to occur when the free charge in the channel goes to zero.

$$Q_I = 0 \tag{2.1.23}$$

then the saturation voltage is:

$$V_{DSAT} = V_{GS} - V_T \tag{2.1.24}$$

The current in the saturation region becomes:

$$I_{DS} = \frac{W}{L}\mu C_{ox}\frac{(V_{GS} - V_T)^2}{1+\delta} \tag{2.1.25}$$

The fact that the mobile charge goes to zero, implies infinite electric field at the pinch-off point. Two dimensional solutions show that the charge density in the channel is never zero and electric field does not become large. At the pinch off point diffusion becomes an important mechanism for transport [Taylor]. After this point current saturation occures and the drift becomes the dominant form of the curent. But the GCA model does not include explicitly current saturation and a mechanism for transfer of curent from the drift mode to the diffusion. As the channel length is reduced and the lateral electric field (and its gradient across the channel) increase, it is expected that these problems with the GCA and the charge sheet model will be agravated.

2.2 Mobility in Inversion layers

Electron mobility in inversion layers is one of the most fundamental parameters in MOS device physics, and must be known for accurate device modeling and design. There is also interest in carrier mobility as an indicator of accuracy of the quantum mechanical models of the two dimensional system at the Si/SiO_2 interface.

In bulk Si, at low temperature ($T < 100$ K), the mobility is limited by the ionized impurity scattering, and the mobility behavior can be very well explained by the Brook-Herring model (At low carrier densities mobilities close to 10^5 $cm^2/V.s$ have been obtained) [Long]. At higher temperture and low carrier densities (below 10^{16}) the bulk mobility is limited by the accoustic phonon scattering . At higher impurity scattering, the ionized impurity scattering again sets in and reduces the mobility.

Mobility behavior in the silicon inversion layer is considerably more complicated. This is due to the two-dimensional nature of the electron gas and the presence of a normal field that confines the carriers next to the surface. Three scattering mechanisms are recognized

as the main causes of mobility degradation in inversion layers: First is the acoustic phonon scattering which is the dominant scattering mechanism for $T > 100$ K and at low carrier densities. Coulumb scattering due to ionized impurity scattering, fixed oxide charges, and interface state charges is the second mechanism which is dominant at low temperature ($T < 100$ K, where the phonon density is low). At room temperature, Coulomb scattering is important for lightly inverted surfaces and low normal fields in the inversion layer. At high inversion levels the contribution of the Coulomb scattering is reduced due to screening of the impurities and charge centers. At very high donor densities, ($> 10^{17}$) impurity scattering becomes important again. The third important carrier scattering mechanism is that of surface roughness scattering. This type of scattering is important under strong inversion.

In silicon inversion layers, mobility μ_{eff} is determined from the drain conductance $g_D = \frac{I_{DS}}{V_{DS}}$:

$$\mu_{eff} = \frac{L}{W}\frac{g_D}{Q_I(V_{GS})}|_{V_D \to 0} \tag{2.2.1}$$

where Q_I can be determined experimentally:

$$Q_I(V_{GS}) = q\int_{-\infty}^{V_{GS}} C_{GC}(V_G)dV_G \tag{2.2.2}$$

and C_{GC} id the gate to channel capacitance.

At room temperature, it is found that the mobility is a strong function of the normal field. According to [Sun and Plummer], as the substrate doping level is increased, the maximum carrier mobility (mobility near the threshold) decreases. This is not a result of impurity scattering, but rather a result of the increase in the effective normal field necessary for inversion. Thus it is found that:

$$\mu_e \propto (E_{eff})^{-\alpha} \tag{2.2.3}$$

where E_{eff} is the effective (normal) electric field and α is an empirically determined constant: $\alpha \approx 0.2 - 0.3$ at low normal fields in which the phonon scattering dominates ($E_{norm} < 0.5\ MeV/cm$) and $\alpha = 1$ at high normal fields where the surface roughness scattering dominates ($E_{norm} > 0.5\ MeV/cm$). This relation seems to be universal, and holds over a wide range of impurity densities. At low carrier concentrations at room temperture, it is found that the universal relation breaks down (probably due to lack of screening).

Effective field in an inversion layer can be written as:

$$E_{eff} = \frac{1}{\epsilon_{SI}}[\frac{1}{2}Q_I + Q_B] \tag{2.2.4}$$

This is the average field felt by the carriers in the inversion layer. Q_B is given by:

$$Q_B = (2q\epsilon_{SI}N_A\phi_B)^{1/2} \tag{2.2.5}$$

and Q_I is given by (2.2.2). As the substrate doping is increased, the threshold increases, and the increased normal field (caused by increased Q_B) at the threshold reduces the carrier mobility. Thus according to this model, in order to increase the carrier mobility, the normal field at the threshold must be decreased. This can be achieved by lowering the substrate doping. Decreasing the substrate doping causes increase in short channel effects in sub-micron channel length devices. In this work a nonuniform substrate doping was used: the channel doping was kept low to minimize the impurity scattering, while just below the surface, the doping was kept as high as possible to reduce the short channel effects.

It has to be emphasized that systematic measurements of mobility in samples in which effect of normal field and impurity scattering were seperated, have not been performed.

2.3 Scaling

As the device size is scaled down, one major aim has been to come up with a set of design rules following which the scaled down devices would operate similar to large devices. Different scaling rules have been proposed: One holds the electric field strength constant as the device is scaled down (constant field scaling). Another set of rules aims for device operation with the constant voltage supply (constant voltage scaling). Another set of rules attempts to operate the scaled device under reduced voltages and higher fields (quasi-constant voltage scaling). Each one of these rules imposes a different set of constraints on the scaling of lengths and doping levels, and results in different trade-offs between speed, power, and power-delay product.

The generalized scaling theory was developed by [Baccarani, 1984]. For a given device geometry, the field configuration within device is given by a solution of the Poisson's and te current continuity equations:

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=\frac{q}{\epsilon_{Si}}(p-n+N_D-N_A) \tag{2.3.1}$$

$$\nabla\cdot\vec{J}=0 \tag{2.3.2}$$

If the spatial dimensions are scaled by $1/\lambda$ and the voltages are scaled by $1/\kappa$ (i.e. the electric field is scaled by κ/λ), and defining the scaled variables:

$$(x',y',z')=(x,y,z)/\lambda \tag{2.3.3a}$$

$$\psi'=\psi/\kappa \tag{2.3.3b}$$

$$(p',n',N_D',N_A')=(p,n,N_D,N_A)\lambda^2/\kappa \tag{2.3.3c}$$

Then the following differential equation is satisfied by the scaled quantities:

$$\frac{\partial^2\psi'}{\partial x'^2}+\frac{\partial^2\psi'}{\partial y'^2}+\frac{\partial^2\psi'}{\partial z'^2}=\frac{q}{\epsilon_{Si}}(p'-n'+N_D'-N_A') \tag{2.3.4}$$

General Scaling Laws

Linear Dimensions:	(W, L, t_{ox}, x_j)	$1/\lambda$
Biases:	V_{GS}, V_{DS}, V_T	$1/\kappa$
Impurity Concentration	N_A, N_D	λ^2/κ
Current:	I_D	λ/κ^2
Electric Field:	E	λ/κ
Transistor Transit Time	t_d	κ/λ^2
Power Delay Product:	$I_D V_D t_D$	$1/\lambda\kappa^2$

Table 2.1

This is identical to equation (2.3.1). If the boundary conditions are scaled too, then the two electric field solutions have identical shape, but are different only in intensity. It must be emphasized that only Poisson equation has been scaled down: The influence of the current continuity equation on the potential distribution has been ignored. This assumption is fine under the subthreshold condition, where the electron concentration is negligible. Under strong inversion, the carrier density in the channel is exponentially dependent on ψ_S, and it no longer scales by λ^2/κ. But if a charge sheet model is assumed (infinitely thin inversion layer), scaling still applies.

Table 2.1 summarizes the scaling factors within the generalized scaling theory. Some of the assertions in terms of speed or power performance of scaling models are questionable. But there is no doubt that scaling improves the packing density and overall speed. In practice, when scaling devices into the deep submicron range, the scaling rules are not followed exactly. The goal is to maximize current drive, mobility, turnoff speed, and minimize output conductance. Some of these goals are contradictory, and depending on the device application, different approaches are taken.

The scaling theory has a number of limitations. The most significant of these is that short channel effects, and turn-off charactristics degradation are ignored. As the device is scaled down , its turnoff characteristics (sub-threshold slopes at different V_{DS}) not only

do not improve, but they worsen. It is no longer possible to scale the threshold (i.e. it is not possible to scale the built-in potentials).

Among other problems with scaling models are the effect of the finite inversion layer thickness. As the gate oxide thickness is reduced below 10 nm, its thickness becomes comparable to that of the inversion layer thickness. Finite inversion layer thickness, reduces the effectice gate drive (from C_{ox}), i.e. the effective gate-channel capacitance is oxide capacitance in series with the bulk and the inversion layer capacitance [Sodini 1982]:

$$C_{GC}(V_{GS}) = \frac{\partial Q_I}{\partial V_{GS}} = \frac{C_{ox}C_{inv}}{C_{ox} + C_{inv} + C_B} \tag{2.3.5}$$

where $C_B = \frac{\partial Q_B}{\partial \psi_S}$ is the bulk capacitance and $C_{inv} = \frac{\partial Q_{inv}}{\partial \psi_S}$ is the inversion layer capacitance. The apparant reduction in the gate drive, is most severe close to the threshold, and causes reduction of I_D. This might look as if the mobility is degraded in the thin oxide device [Han], but if the C_{GC} is measured properly, no degradation is observed [Liang].

In scaled devices there is mobility degradation due to increase in the substrate doping and the resultant increase in the normal field under which the device operates. This effect is not included in a straight forward scaling model. As devices are scaled down they operate under larger lateral fields too. This causes velocity saturation.

In scaled devices, the parasitic resistances per device increase. The amount of degradation caused by the parasitics becomes more important as the devices operate under larger current densities.

Despite all these problems, scaling has provided general guidelines for design and understanding of small device, and works very well in the in the $L > 0.5$ μm range.

2.4 Short Channel Effects

As the MOSFET channel length is scaled down into the deep submicron range, un-

wanted effects associated with the drain control of the channel and the two dimensional structure of the MOSFET, become more prominent. Among these are reduction of the threshold as the channel length is reduced, the increased output conductance, reduction of the subthreshold swing (slope), reduction in body effect coefficient, and dependence of threshold voltage on V_{DS} (drain induced barrier lowering or DIBL). These effects are closely related, are caused by sharing of the space charge layer of the drain-bulk junction and the channel depletion charge at the drain side of the channel. When this common charge becomes a significant part of the channel charge, then the drain voltage influences the channel properties. The charge sharing model, used in modeling of short channel devices is presented next. There are a number of other effects that take place mainly in the short channel devices (hot-electron effects and velocity saturation) and have a different origin. They are explained later.

2.4.1 Charge Sharing Model

This model concerns itself with the effect of drain and source depletion layer on the threshold of a short channel device. This is a very simple conceptual model and it is not possible to justify all its steps.

Consider the short channel MOSFET in Fig. 2.4. When calculating the threshold in a long channel device, it is assumed that all the charges in the inversion layer and the depletion layer have their image on the gate electrode, i.e. this is essentially a one dimensional problem. In other words, the gate bias must be large enough to support all the charges in the channel. In a short chanel device, a significant fraction of the field lines from the channel charges terminate on the drain. Thus, the extra charge on the gate (at a given gate bias), must terminate on the extra inversion charges in the channel, i.e. more

I_D and lower V_T. In a long channel device the threshold voltage is:

$$V_T = V_{FB} + \phi_B - \frac{Q_B}{C_{ox}} \tag{2.4.1}$$

If the channel is short, the fraction of the channel whose field lines terminate on the drain, has to be subtraced from the total depletion charge in the threshold formula ($Q_B L$) (Figure 2.4):

$$V_T = V_{FB} + \phi_B - \frac{Q_B L - \Delta Q_B}{LC_{ox}} \tag{2.4.2}$$

Thus the reduction in the threshold voltage due to charge sharing is:

$$\Delta V_T = \frac{\Delta Q_B}{LC_{ox}} \tag{2.4.3}$$

For a n^+ junction of radius r_j, and space charge layer thickness l_S at $V_{DS} = 0$, using simple geometrical argument:

$$\Delta V_T = \gamma \frac{r_j}{L}(\sqrt{1 + \frac{2l_S}{r_j}} - 1)\sqrt{V_{SB} + \phi_B} \tag{2.4.4}$$

For a deep junction, a Taylor expansion of the above formula can be obtained:

$$\Delta V_T = \frac{qN_A W_{max}}{C_{ox}} \frac{l_S}{L} \tag{2.4.5}$$

This expression states (as expected) that the charge shared is proportional to the product of space charge layer thickness and the substrate doping:

$$\Delta Q_B \propto N_A \times l_S \tag{2.4.6}$$

The fraction of the charge shared between channel and drain junction is:

$$\frac{\Delta Q_B}{Q_B} \propto \frac{l_s}{L} \propto \frac{1}{L}\sqrt{\frac{1}{N_A}} \tag{2.4.7}$$

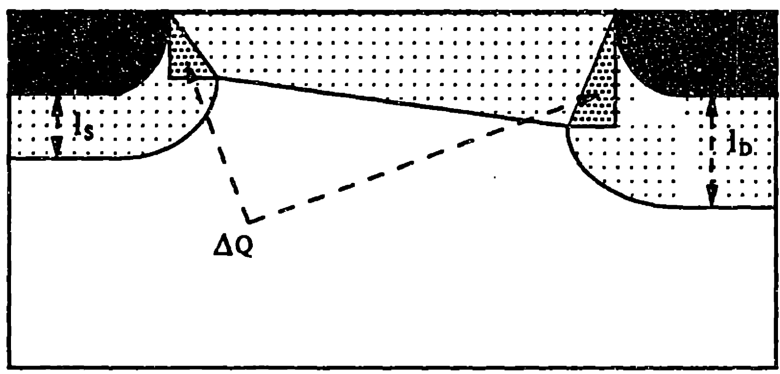


Figure 2.4: Cross section of a short channel MOSFET, showing the the charge shared, ΔQ_B

Expression (2.4.7) shows why increasing the substrate doping reduces the short channel effects. Increase in substrate doping increases the net amount of the charge shared, but reduces the fraction of the charge shared between n^+ junctions and the channel.

Putting a heavily doped substrate just below the surface would also reduce the short channel effect (while keeping the threshold constant): The heavy p^+ doping just below the surface, reduces the depletion layer width just below the surface. Then the charge shared is the product of a small $l_S \propto \sqrt{1/N_A^+}$ and a light surface doping $\propto N_A$, and the charge shared is reduced (Figure 7.2, Chapter 7).

If there is finite bias on the drain and source, ΔV_T can be approximated (for small V_{DS}) as:

$$\Delta V_T = \frac{qN_A W_{max}}{C_{ox}}(\frac{l_S + l_D}{2L}) \qquad (2.4.8)$$

where l_S and l_D are the space charge layer thicknesses at finite source and drain biases:

$$l_D = \sqrt{\frac{2\epsilon}{qN_A}}\sqrt{\phi_B + V_{DB}} \qquad (2.4.9)$$

For small V_{DB} (compared to $2\psi_F$), threshold shift is:

$$\Delta V_T = \alpha_1 \frac{\epsilon_{Si}}{\epsilon_{ox}} \frac{t_{ox}}{L}[(\phi_B + V_{SB}) + \alpha_2 V_{DS}] \qquad (2.4.10)$$

This relation predictes a linear variation of ΔV_T. Although this was derived for small V_{DB}, measurements and two dimensional simulation shows that this relation holds for large values of V_{DB} also (Figure 7.2). α_1 and α_2 are constants that depend on the process and channel length. According to this relation, reduction in t_{ox} would reduce the short channel effects.

The charge sharing model is an intuitive tool for understanding short channel effects. It does not address issues raised by the breakdown of the gradual channel approximation.

When designing short channel MOSFETs, one must use two dimensional simulators for accurate predication and models.

2.5 Velocity Saturation

As the magnitude of the lateral electric field applied to the semiconductor increases, the electrons begin to gain energy. To balance the energy gain from the field, there is an increase in the rate of energy loss to the lattice. The primary loss mechanism at high fields, is optical phonon emission. At very high fields, because of the increase in the phonon emission rate, velocity saturates. The high field transport, and saturation velocity in the steady state, are governed primarily by the interaction of the electrons and phonons, and are independent of impurity scattering and the details of low field mobility.

The saturation velocity v_s is given as:

$$v_s = \sqrt{\frac{3h\omega_o}{4m}\tanh\frac{h\omega_o}{2kT_o}} \tag{2.5.1}$$

where $h\omega_o$ is the energy of optical phonon emission, and T_o is the lattice temperature. When used in device modeling, it is useful to obtain a relationship between mobility and electric field that includes the velocity saturation term also. One such relationship is:

$$\mu(E) = \frac{\mu_o}{[1 + (\frac{E}{E_C})^\beta]^{1/\beta}} \tag{2.5.2}$$

where E_C is the critical field for the velocity saturation, v_s:

$$E_C = \frac{v_s}{\mu_o} \tag{2.5.3}$$

For silicon, β values between 1 and 2 have been used.

2.5.1 MOSFET with Velocity Saturation

In this section a simple theory of MOSFET under the effect of strong lateral field and velocity saturation is presented. This model is useful in explaining some of the effects seen in short channel MOSFETs. The mobility model used is as follows:

$$v = \frac{\mu_{eff}E}{1 + E/E_c} \qquad E < E_c \tag{2.5.4}$$

$$= v_{sat} \qquad E > E_c$$

where E is the lateral field and $E_c = 2v_{SAT}/\mu_{eff}$ is the critical field for velocity saturation. μ_{eff} is the effective carrier mobility:

$$\mu_{eff} = \frac{\mu_o}{1 + a(V_{GS} - V_T)} \tag{2.5.5}$$

where a ia an experimentaly determined constant. Furthermore in this model it is assumed that the drift dominates the total current, and the total number of the carriers in the channel is given by $C_{ox}(V_{GS} - V_T)$. In the linear region, the current at any point along the channel is the product of the carrier velocity, $v(x)$, and density:

$$I_D = I(x) = WC_{ox}(V_{GS} - V_T - V(x))v(x) \tag{2.5.6}$$

Expressing $v(x)$ in terms of $E(x) = -\frac{dV(x)}{dx}$, and integrating from source to drain, one obtaines:

$$I_D = \frac{W\mu_{eff}C_{ox}(V_{GS} - V_T - V_{DS}/2)}{L(1 + V_{DS}/E_cL)}V_{DS} \tag{2.5.7}$$

this is valid for operation in the saturation region. Saturation of current is at the point where carriers are moving with the saturation velocity, i.e. $E = E_c$ at the drain side. At the saturation point:

$$I_D = Wv_{SAT}C_{ox}(V_{GS} - V_T - V_{DSAT}) \tag{2.5.8}$$

where v_{SAT} is the drain saturation velocity. Solving for V_{DSAT} by equation Eq. (2.5.8) and Eq. (2.5.7), one obtaines:

$$V_{DSAT} = \frac{E_c L(V_{GS} - V_T)}{E_c L + (V_G - V_T)} \tag{2.5.9}$$

As one scales the channel length, it takes less drain bias for the critical field to get to E_c, and thus V_{DSAT} is reduced from its first order theory. This simple model will later be used in chapters on velocity overshoot and hot electron effects.

Chapter 3

Fabrication

This chapter reviews the techniques used in the fabrication of deep submicron MOSFET. The basic process was developed by [Chou, 1986]. Some improvements in his process have been made.

3.1: Introduction

As the device dimensions are reduced and the density of devices is increased, a number of new fabrication obstacles are encountered. It is no longer possible to use techniques used in fabrication of deep long channels and apply them to fabrication of submicron devices. The foremost among them is the issue of lithography. Because of the limits of optical lithography it is not possible to reliably put down linewidths below 0.5 μm using commercially available G-line or H-line optical steppers (Using high NA oil immersion lenses, Kausta et al have achieved 0.13 μm linewidth, but this was a research demonstration). This is not the only the limitation of optical systems: the variations in the linewidths defined with optical lithography become unacceptable in the submicron range. That is why one has to resort to shorter wavelength lithographic tools e.g. X-ray or electron beam.

Lithography is not the only technological issue faced in fabricating of short channel devices: Because of scaling of all device dimensions, e.g. the junctions, oxide thickness and

the contact areas, etching and patterning tolerances of most fabrication steps decrease. More controlable, uniform, and selective fabrication steps have to be developed. Probably the biggest challenge in utilization of deep submicron MOS devices will be the reduction of mismatch among devices and uniformity across wafer. As the dimensions are reduced the devices operate at higher speeds and current densities. It is extremely important (and very difficult because of shallower junctions) to reduce parasitic resistances and capacitances. Small changes in parasitics will cause unacceptable variations in performance.

At MIT we did not have access to e-beam lithogrophy, and there was a x-ray exposure system available. A process based on use of this system to fabricate short channel devices was developed here by Chou. Since the details of his process were reported earlier, the details will be only briefly repeated here, along with the modifications made in the current thesis work. First the fabrication of x-ray mask would be reported, and then the details of fabrication steps are presented. The results will be compared with those obtained earlier.

3.2 Fabrication of X-ray Mask

To fabricate short channel MOSFET's, narrow linewidths must be defined. The conventional patterning tool for small linewidths has been either electron-beam lithography or x-ray lithography. Previous work in the Submicron Structures Lab (SSL) at MIT has made use of x-ray lithography. The x-ray mask membrane has been organic (polyimide). Thus, there was significant distortion across the mask. As a result one is not able to align the mask to a previous layer. This excludes use of the mask for self-aligened structures (This problem would go away if one used inorganic masks, a project currently being pursued in the SSL). As a result, the x-ray mask must be used as the first layer and all the subsequent layers must be aligned to it. Since we are fabricating short channel MOSFET's, x-ray will be used to define source and drain regions in x-ray resist, the PMMA. The pattern

in PMMA will later be used as a mask against ion implantation, during the formation of source and drain regions.

X-ray masks memebranes used are thin polyimide ($\approx 1\mu m$ thick) stretched on a vespel ring, as shown in Fig. 3.1. Coarse patterns are defined on the back of the mask by means of optical lithography. Fine features are defined in gold on the front of the mask by means of shadowing to a mesa step formed in polyimide. A good part of fabrication of the x-ray mask involves forming vertical sidewall mesas on Polyimide for the susequent shadowing.

X-ray masks are fabricated as follows: (110) Si wafers (with flat along (111)) are oxidized in dry O_2 (t_{ox} = 450 A). This oxide is later used as a mask for formation of mesas. On two corners of the wafer, sets of alignment mark, each rotated 0.5^o with respect to the other, are placed along with rectangles. The goal here is to find a set of alignment marks which correspond to well that are optimally aligned to the crystallographic axis. Oxide is etched below this pattern. Resist is striped. Using anisotropic etching ($KOH : H_2O : Isopropylalchohal$ 3 : 8 : 2 by volume at 80^oC) Si is etched (0.2 μm for about 1 minute). Etch rate is about 100 times faster on (110) plane (bottom of the wells) as compared to the (111) planes (sides of the wells). If we are very closely aliegned to the (111) plane, very smoth well walls are obtained (Figure 3.2).

Also, along the intersection of (111) planes and (110) planes the etch rate is different and as a result striations appear on the bottom of the square wells. These lines make an angle of 35.3^o with the (111) plane. Therefore, one has to find the rectangle in which the striations on the bottom of well make an angle of 35.3^o with the wall of the well. The wall of that well is along (111) and the corresponding set of alignment marks are exactly parallel to the (111) plane. Close examinations of the walls that are off alignment by even 2°, indicates that any steps in these walls are less than 10 nm.

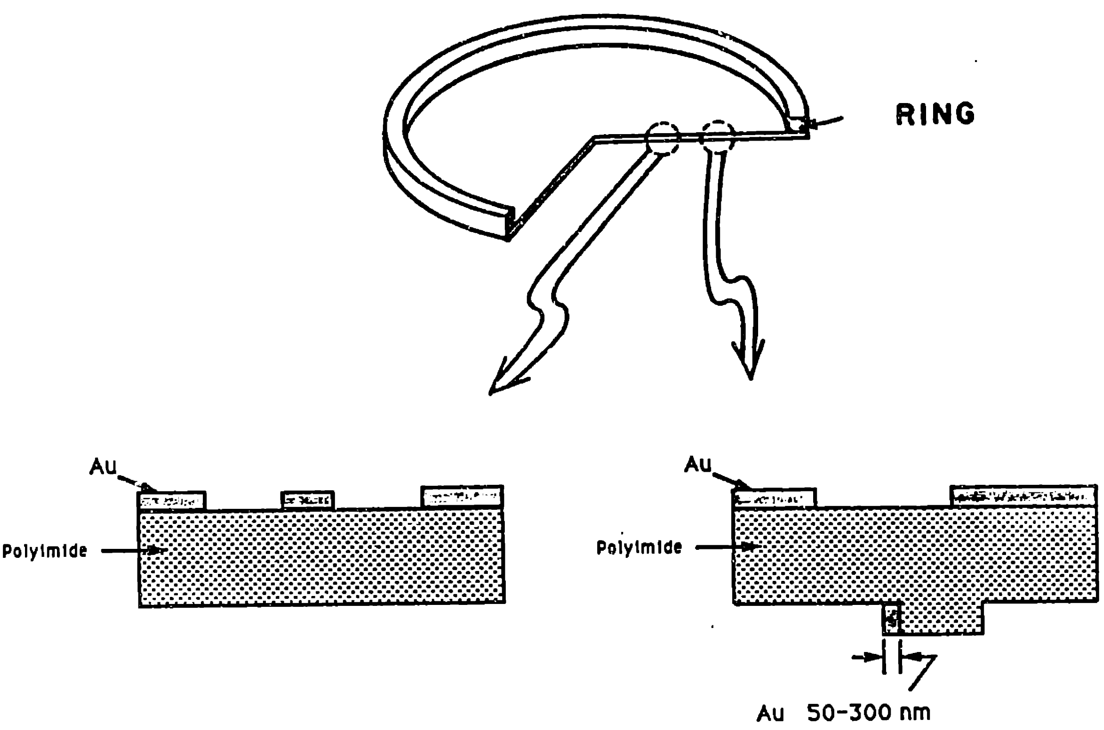


Figure 3.1: An x-ray mask: A thin polyimide memberane stretched on a vesple ring

Anisotropic Etching of Si

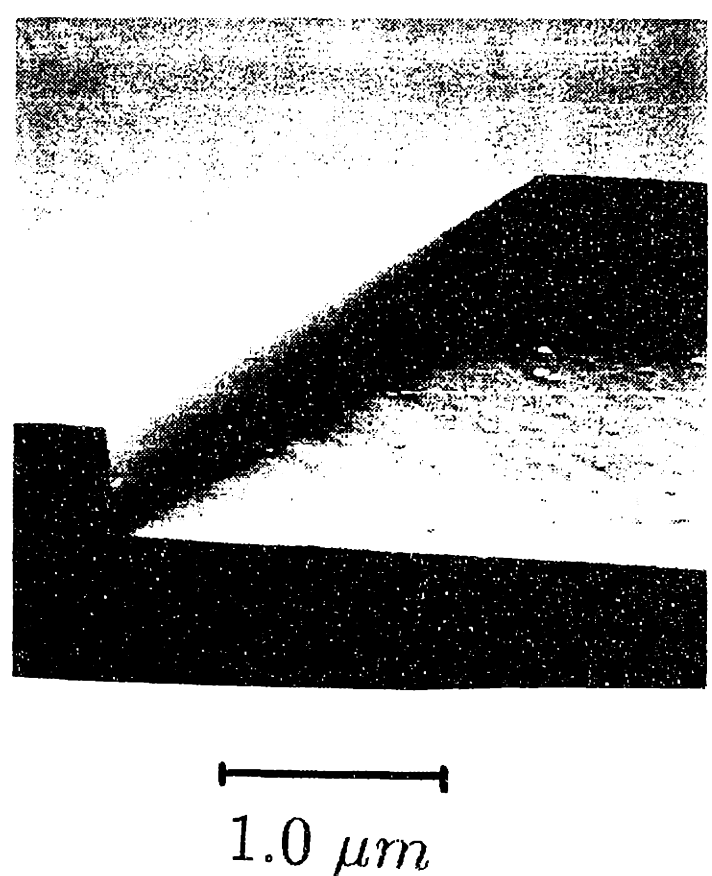


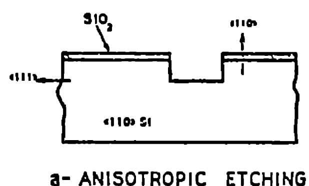
Figure 3.2 SEM micrograph of a (110) Si wafer in which wells were etched. The side wall to the left is (111) plane. Note the extremely smooth sidewall that results from anisotropic etching.

After finding the right alignment mark (the one that is aligned to the (111) plane), the mesa pattern is defined by optical projection lithography, and etched into oxide all over the wafer. Next, silicon is etched in the same anisothropic etch as before for about 0.25 μm (Figure 3.3a). Then oxide is stripped everywhere and about 1 μm of Polyimide (Dupont, type PI-2555) is spun on . After curing at 80°C for 30 minutes and a 150°C to 250°C ramp in about 3 hours, it is cured at 400° for 30 minutes (Figure 3.3b). About 10 nm of chromium (for adhesion) and 150 nm of gold is e-beam evaporated. Next the coarse pattern (on the back of x-ray mask) is defined in photoresist (all features of the mask, except the submicron lines). Gold is patterned in an ion miller (Wet etch although easier causes severe undercutting and loss of critical dimension) (Figure 3.3c). Then, a espel ring (about 1 inch in diameter) is attached and silicon is etched from the back in $HF : HNO_3$, 97 : 3. If there is too much bubling, the amount of HNO_3 can be reduced further. After the etch, the thin membrane is left streched on the ring. Then the fine line gold absorber is produced by angle evaporation on mesa side walls at 4° to the desired thickness (Figure 3.3d).

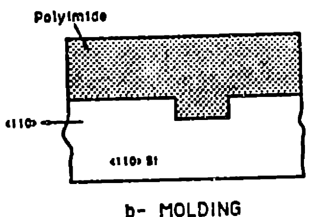
In most of our work, 150, 200 1nd 250 nm absorber lines were used. Previously the excess gold was ion milled from the back surface. It was found that there is no need for this since the very thin layer of gold on back is transparent to x-ray. Moreover, ion milling causes excess heating of the mask (and breakage). As a last step, 50 nm of Al is evaporated on the back of the mask to help in forming contact between the mask and the wafer during the x-ray exposure. The mask is used to expose about 0.2 μm of PMMA on device wafers.

3.2.2 X-ray Exposure

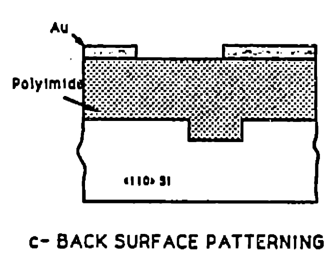
The x-ray mask is used to carry out x-ray exposure. The x-ray source is the C_k line of a graphite target (λ =4.5 nm). The mask was held intimately into contact with the



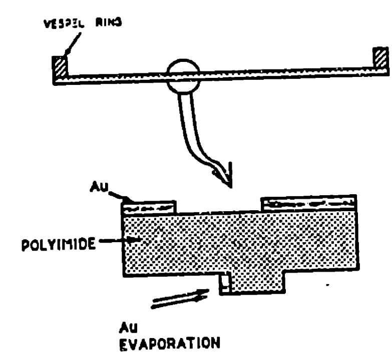
a- ANISOTROPIC ETCHING



b- MOLDING



c- BACK SURFACE PATTERNING



d- MOUNT VESPEL RING
ETCH Si
SHADOW EVAPORATE Au

Figure 3.3: Steps involved in fabrication of x-ray mask

substrate. The substrate is a lightly doped p-type ($\rho =$ 10-20 $\Omega - mm$), coated with 220 nm of PMMA. The distance beween the source and the mask is about 10.5 cm, and for our low power source (300 Watt), exposure takes about 12-18 hours. After exposure the samples were developed in isopropyl alchhol-MIBK solution (ratio of 60-40 by volume). A typical line obtained by the x-ray exposure is presented in Fig. 4.4.

3.2 Device Fabrication

The pattern obtained by x-ray lithography is used as a mask for the source and drain implantation. Arsenic, with an energy of 30 KeV and a dose of 7×10^{15} per cm^2 is used for n^+ source and drain. The 200 nm PMMA thickness is more than enough to stop the As, whose range is $R =$ 31 nm and straggle is $\Delta R =$ 9 nm. To stop 99.999% of the As, the thickness of PMMA must be greater than $R + 4.3\Delta R$ =70 nm.

After the x-ray exposure, one should lay down alignment patterns for later steps. (The distance between alignment patterns is 76.2 mm.) To do this one has to correct for the distortion in the x-ray mask, i.e. find the new stepping distance. The distortion occurs after the release of the memberane from the Si due to the plastic nature of the polyimide. After implantation of source and drain, the wafer is coated with 1 μm of resist. Using the laser interferometer of the stepper, the distance between the alignment marks on the x-ray exposed pattern is found. In Fig. 3.5, the measured relative motion of the alignment patterns with respect to the center of the mask, is presented. As expected, the the alignment marks at the edge show little movement. As it can be seen it is hard to come up with a new steeping distance to cover all the dies.

The rest of the fabrication is standard, yielding non-self-aligned, metal gate MOSFETs. After the alignment marks for later steps were defined on the wafer (Mask 1), they are etched in the silicon, off the side of the PMMA. The resist and the PMMA

X-ray Exposed Resist for Ion Implantation of Source and Drain

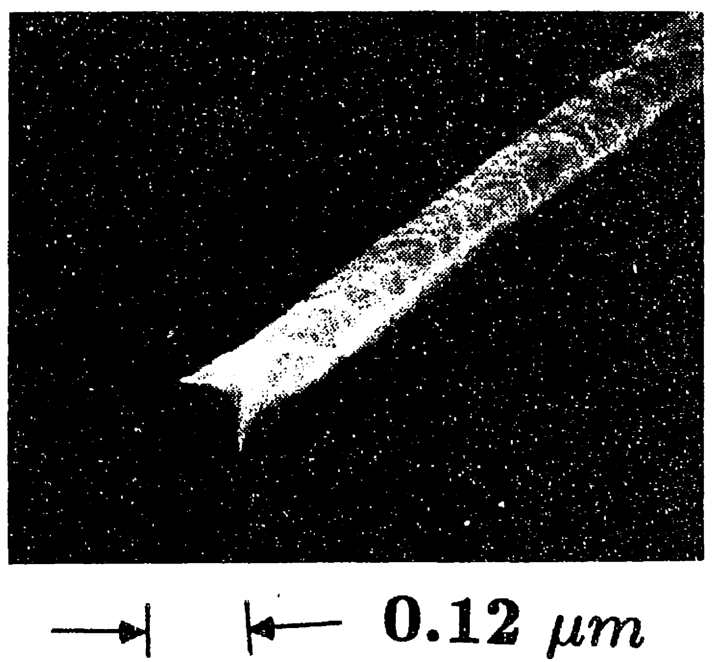
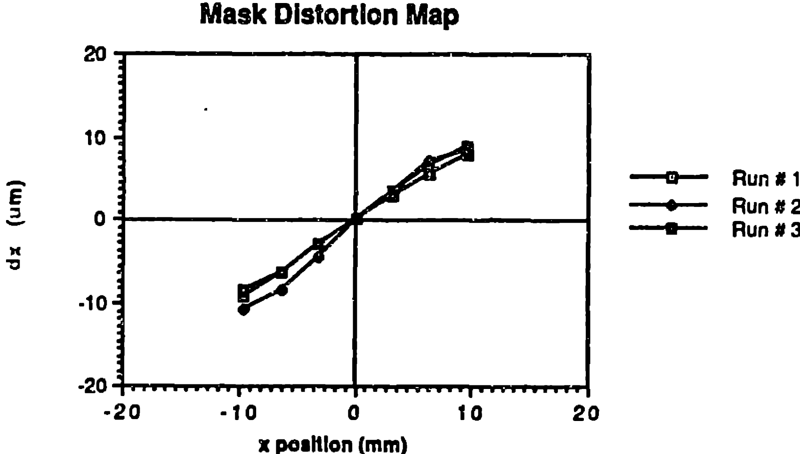


Figure 3.4: A 0.12 μm wide line obtained by x-ray exposure



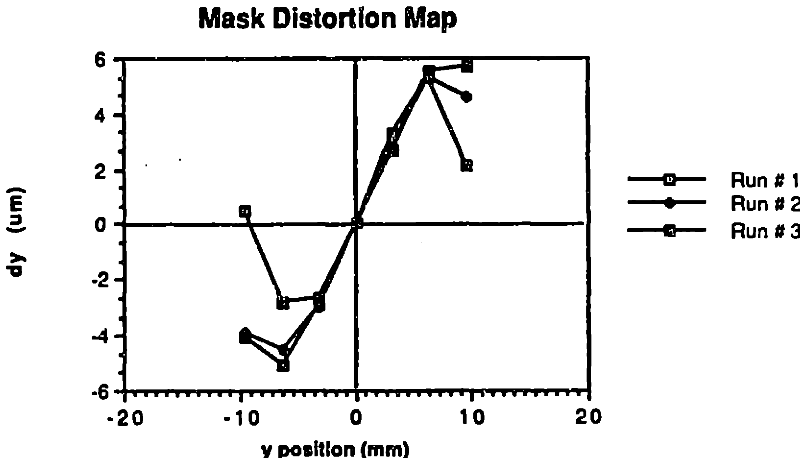


Figure 3.5: Movement of dies (relative to the center of x-ray mask), as measured by the stepper.

Improvements in Process Results

	Previous Results	These Results
Minimum t_{ox}	11 nm	2.5-7.3 nm
Shallow n^+ resistivity	140Ω/	90Ω/
Deep n^+		30 Ω/
R_{SD}	7000 $\Omega - \mu m$	600-1400 $\Omega - \mu m$
μ_e at 300 K	300 $cm^2/V.s$	350-450 $cm^2/V.s$
Measured g_m at 300 K	140 mS/mm	upto 710 mS/mm

Table 3.1

are stripped next. The phosphorous deep implant for source-drain contacts (Dose=3E15, Energy=160 KeV) is next (Mask 2). Isolation implantation is next (Boron, Dose=1E13, Energy=30KeV) (Mask 3). The deep punchthrough implant (Dose=1E13, Energy=180 KeV) and the channel implant are next. The implants are annealed at 1050 ^oC for 5 seconds in a rapid thermal annealer in an oxygen ambient. The gate oxide is grown next. For gate metal, Al-Cr-Al (200nm-10nm-200nm) is evaporated and patterned (The chromium layer is used as etch stop of PAD etch when openning the contact hole to gate) (Mask 4). CVD oxide is put down next and patterned for the contact hole, and etched for the contacts (Mask 5). Al-Cu-Si is sputtered and paterned everywhere (Mask 6). Finally the devices are sintered at 400° C for 5 minutes.

3.4 Results

In this section a summary of the results of the fabrication process are presented, First the issues with regard to source and drain resistivity are presented. Next a summary of the issues relevant to gate dielectric is given.

The yield depends on the design rule and the amount of overlap between the metal gate and the channel. A design rule of 2.5 μm was used. On good runs the yield was close to 90%.

3.4.1 Contact and Sheet Resistivities

To calculate the carrier velocity, the saturated transconductance of the device, g_m, is measured. After the necessary corrections for the source and drain resistance, R_{SD}, the intrinsic transconductance, g_{mi}, is obtained. g_{mi} is directly proportional to the average electron velocity in the channel. To minimize corrections due to R_{SD}, the source-drain resistance must be small. To reduce R_{SD}, a number of modifications to the previous process were used: tighter design rules were used than in earlier work, and the source and drain contact were brought to within 5 μm of channel edge. To further reduce the source and drain resistances, rapid thermal annealing was used to increase the activation of the n^+ layers. R_{SD} was reduced to 1/10 of the value obtained in the previous work.

4.4.2 Gate Oxide

To minimize the short channel effects, the oxide thickness was also reduced (Down to 2.5 nm in some samples). Thin oxides were grown at low temperature in pure oxygen ambients. The oxydation times were short (From 5 minutes at 800^oC for 2.5nm to 12 minutes at 900^oC for 7.5 nm). It was found that the quality of the oxide, in terms of amount of tunneling current, were critically dependent on cleanliness.

3.4.3 Process Result Summary

In this section a comparision of the process results is made with the previous results [Chou]. This process, when all the steps worked and no complications were encountered, produced remarkably good devices. In Fig. 3.6, the device charactristics for a 0.15 μm device is presented. This device has a measured transconductance of 710 mS/mm and a intrinsic tranasconductance of more than 1000 mS/mm, which is the highest reported to

date. This has been made possible by high channel mobility and thin oxide. Table 3.1 summarizes the improvements in this process as compared to [Chou, 1986].

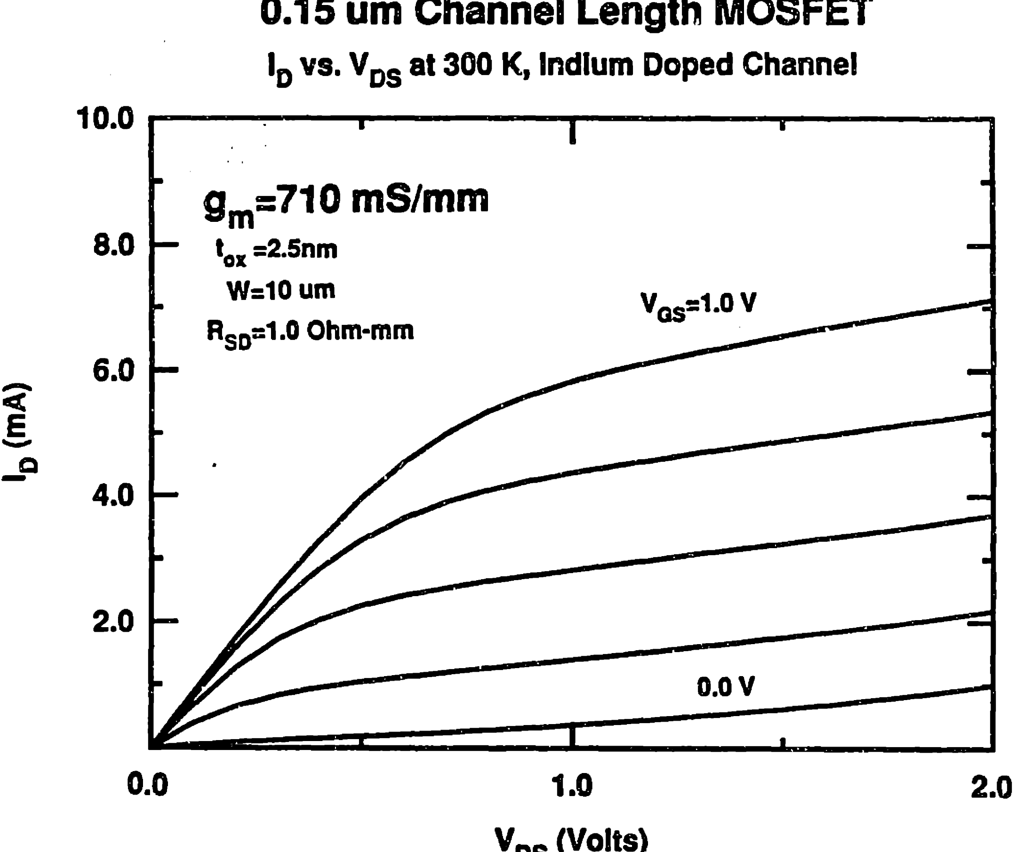


Figure 3.6: A 0.15 μm device produced by our process. This device has the highest g_m reported to date, despite its relatively high R_{SD}.

Chapter 4

Non-Stationary Transport

Theoretical work in device physics and modeling are concerned with detailed analysis of collision process, scattering probabilities, determination of distribution functions, velocity versus field characteristics, and ultimately current versus voltage characteristics. Under static conditions, in long (homogeneous fields) devices, one can make use of a priori displaced Maxwellian distribution, or more ambitiously determine the distribution function. In either case the carrier velocity is obtained from drift-diffusion (DD) equation:

$$\vec{v} = -\mu(E)\vec{E} - D(E)\frac{\nabla n}{n}$$

and the current density is given by $\vec{J} = -qn\vec{v}$. In DD equation is obtained from the first moment of Boltzmann Equation (Ignoring quantities obtained from the higher oder moments), and mobility and diffusion constants are functions of the local electric field, i.e. the effects due to rapid variations in electric field and energy (higher orders of BE) are iognored. In this chapter a review of classical transport, its limitation in small devices, and the Boltzmann equation are presented first. Then the hydrodynamic transport models as developed by Stratton , Blotekjaer , and Baccarani and Wordeman are presented. They are applied to

one dimensional systems, and the effect of carrier mobility on non-stationary effects (velocity overshoot) is studied.

4.1 Non-Stationary Transport

Rapid developements in very large scale integration and introduction of submicron devices in MOS circuits has raised questions about the validity and suffuiciency of drift-diffusion models. In commonly used DD models, it is assumed that the mobility and diffusitivity are functions of local electric field. This is a valid assumption as long as the electric field varies slowly: Carrier temperature is always in equilibrium with the electric field. But in small devices devices, electric field and carrier density vary significantly across the device. Carrier temperature is out of stationary equilibrium with the local electric field. Questions about the validity of simple DD equations are raised. Certainly DD is unable to predict some of the new phenomena predicted by the use of Monte Carlo solution to Boltzmann Equation (BE). Therefore any model based on DD is incapable of predicting the device characteristics where electric field and carrier density vary very rapidly in space and/or time.

To remedy this, and model the new phenomena observed in small devices, a number of techniques have been devised. The first one is the modification of DD equations by Thornton (and followed upon by M. Artaki and P. J. Price). In this technique carrier velocity in DD has been modified.

$$\vec{v} = -\mu(E)\vec{E} - \mu(E)\vec{E} \times L(E)\frac{\nabla E}{E} - D(E)\frac{\nabla n}{n} \tag{4.1.1}$$

where $L(E)$ is a phenomenological length coefficient:

$$L(E) \approx \frac{1}{q}\frac{d < W >}{dE}$$

where $< W >$ is the average carrier energy. $L(E)$ has been introduced to take care of some of the phenomena caused by the rapid variation in electric field, and can be calculated either heurestically or by use of the Monte Carlo technique. So far this modified current equation has not been implemented in any device modeling explicitly.

Another technique is the so called Hydrodynamic model, based on calculating the higher moments of BE. In this technique, mobility and diffusion coefficients become functions of carrier temperature (as opposed to functions of electric field as done in simple DD). Moments of BE provide the equations for the carrier temperature. This technique will be described in detail in next section.

Another technique for modeling transport in small devices is use of Monte Carlo technique. This is rigorous, but takes a lot of computer time to get to a self-consistant result. Details of this technique can be found in [Jacoboni].

Of all these techniques, the hydrodynamic model is probably the most promising, especially when the goal is device design. This technique can be based on semi-emperical data, and has shown remarkable ability not only to model carrier veleocity overshoot, but in two dimensional simulators it seems to be capable of showing the reduction in hot electron effects in small devices.

4.2 The Boltzmann Equation

To describe the complete electronic system, the distribution function, $f(\vec{r}, \vec{k}, t)$, which gives occupation probability for a state charactrized by k-vector $\vec{k}$, and space vector $\vec{r}$, is introduced. For a homogeneous solid in equilibrium, $f(\vec{r}, \vec{k}, t)$ is the Fermi distribution function. To calculate the distribution function, given external field, we examine its behavior with time. If there is no scattering from (k, dk) to (k', dk'), then $df/dt = 0$. But if we consider df/dt to be due to collisions only, then:

$$\frac{df(\vec{r}(t), \vec{k}(t))}{dt} = \frac{\partial f}{\partial t} + \dot{\vec{k}}.\vec{\nabla}_k f + \dot{\vec{r}}.\vec{\nabla}_r = \frac{\partial f}{\partial t}|_{collision} \tag{4.2.1}$$

where:

$$\dot{\vec{k}} = -\frac{e}{\hbar}(\vec{E} + \vec{v} \times \vec{B}) \tag{4.2.2}$$

$$\dot{\vec{r}} = \frac{1}{\hbar}\nabla_k(E(k)) \tag{4.2.3}$$

and:

$$\frac{\partial f}{\partial t}|_{coll} = \sum_{k'}[\mathcal{P}(k, k')f(k')(1 - f(k)) - \mathcal{P}(k', k)f(k)(1 - f(k'))] \tag{4.2.4}$$

where: $\mathcal{P}(\vec{k}, \vec{k}')$=Transition probability from $\vec{k}$ to $\vec{k}'$ and:

$$f_o(E_k) = \frac{1}{1 + e^{(E_k - E_F)/kT}} \tag{4.2.5}$$

where E_F is the Fermi energy and k is the Boltzmann constant. In DD models, the relaxation time approximation is used, i.e. a scattering time is introduced.

$$\frac{\partial f}{\partial t}|_{coll} = -\frac{f - f_o}{\tau} = \frac{f_1}{\tau} \tag{4.2.6}$$

It can be shown that this leads to the common local DD transport equations [Ziman]:

$$f(\vec{r},\vec{k},t) = f_o(\vec{r},\vec{k} - \frac{q\tau}{\hbar}\vec{E},t) \tag{4.2.7}$$

i.e. carriers are moving with the drift velocity $\frac{q\tau}{m}\vec{E}$, were q is the carrier charge. In this formulation, the carier velocity is a local function of electric field and all dependencies on non-uniformaties in electric field vanish.

4.3 The Hydrodynamic Model

The hydrodynamic set of transport equations is derived from the higher moments of BE. Within this formulation, carrier mobility is a function of carrier temperature. Equations governing the carrier temperature must be developed. They are a first attempt to investigate the effect of non-uniformities in electric field and doping and time evolutions in transport. They form a set of coupled drift-diffusion and energy balance equation.

Most fundamental, and questionable, assumption used in hydrodynamic equations is that of using a displaced Maxwellian: It is assumed that the carrier distribution at all times is a displaced Maxwellian.

$$f_o(\vec{x},\vec{u},t) = \hbar^3(2\pi mkT)^{-1.5}e^{-\frac{m(\vec{u}-\vec{v})^2/2}{kT}}$$

where $\vec{u}$ is the particle velocity and $\vec{v}$ is the group (average) velocity. The only justification for using a displaced Maxwellian is that the electron-electron scattering time (The mechnism responsible for pushing the distribution toward its equilibrium) is much shorter than all the other pertinent times in the calculations. This asumption is certainly not true in GaAs, as

numerous Monte Carlo simulations in hot electron transistors have shown. In silicon this assumption migth be more justified.

4.3.2 Hydrodynamic Equations

In this section full hydrodynamic set of transport equations are presented. First their deivation is reviewed, and then they are solved in one dimension by a different method to that of previous works. These equations were first obtained in their complete form by Blotekjaer. Blotekjaer removed the simplifing assumptions used by Stratton and others and derived the general equations for velocity and energy that includes thermal conduction for a multi-vally semiconductor. Baccarani and Woordeman applied the the set of equations for multivally semiconductor to silicon. They starts with the BE and integrates it over all space:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial \vec{r}}.(n\vec{v}) = \frac{\partial n}{\partial t}|_{coll} \tag{4.3.1}$$

This is the charge conservation equation. The momentum conservation equation, or the general DD equation, obtained by taking the first moment ofthe BE is:

$$\frac{\partial \vec{P}}{\partial t} + \frac{\partial}{\partial \vec{r}}.(n\vec{v})\frac{\vec{P}}{n} + (n\vec{v}).\frac{\partial}{\partial \vec{r}}\frac{\vec{P}}{n} + qn\vec{E} + \frac{\partial}{\partial \vec{r}}nk\vec{T} = \frac{\partial \vec{P}}{\partial t}|_{coll} \tag{4.3.2}$$

The general energy conservation equation is:

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial \vec{r}}.(W\vec{v})\frac{\vec{P}}{n} + \frac{\partial}{\partial \vec{r}}nk\vec{T}\vec{v} + qn\vec{E}.\vec{v} + \frac{\partial}{\partial \vec{r}}.(\vec{Q}) = \frac{\partial W}{\partial t}|_{coll} \tag{4.3.3}$$

Where:

$$n = \int f(\vec{r}, \vec{u}, t)d^3k \tag{4.3.4}$$

is the carrier density per unit area, and

$$\vec{P} = m^* \int \vec{u} f(\vec{r}, \vec{u}, t) d^3 k \tag{4.3.5}$$

is the momentum density carried by the carriers, and

$$nkT_{ij} = m^* \int (u_i - v_i)(u_j - v_j) f(\vec{r}, \vec{u}, t) d^3 k \tag{4.3.6}$$

is the temperature tensor.

$$W = \frac{m^*}{2} \int u^2 f(\vec{r}, \vec{u}, t) d^3 k \tag{4.3.7}$$

is the electron energy density. $\vec{Q}$ is the heat flux:

$$Q_i = \frac{m^*}{2} \int (u_i - v_i)(\vec{u} - \vec{v})^2 f(\vec{r}, \vec{u}, t) d^3 k \tag{4.3.8}$$

Assuming the temperature is a scalar, i.e. $\vec{T} = T\vec{I}$, then the heat flux can be written as:

$$\vec{Q} = -\kappa \frac{\partial T}{\partial \vec{r}} \tag{4.3.9}$$

where κ is the electron thermal conductivity given by the Wiedemann-Franz Law:

$$\kappa = (2.5 + r)(\frac{k}{q})^2 \sigma T \tag{4.3.10}$$

The scattering rates are due both to generation recombination, $\frac{\partial n}{\partial t}|_{coll}$, and intra-band collision, (ibc), i.e.

$$\frac{\partial \vec{P}}{\partial t}|_{coll} = \frac{\partial \vec{P}}{\partial t}|_{ibc} + \frac{\vec{P}}{n} \frac{\partial n}{\partial t}|_{coll} \tag{4.3.11}$$

$$\frac{\partial W}{\partial t}|_{coll} = \frac{\partial W}{\partial t}|_{ibc} + \frac{W}{n} \frac{\partial n}{\partial t}|_{coll} \tag{4.3.12}$$

Ignoring generation-recombination, i.e.

$$\frac{\partial n}{\partial t}\Big|_{coll} = 0 \tag{4.3.13}$$

Introducing the average momentum $\vec{p} = \frac{\vec{P}}{n}$ and average energy $w = \frac{W}{n}$, then the momentum and energy conservation equations become:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial \vec{r}}.(n\vec{v}) = 0 \tag{4.3.14}$$

$$\frac{\partial \vec{p}}{\partial t} + (\vec{v}.\frac{\partial}{\partial \vec{r}})p + q\vec{E} + \frac{1}{n}\frac{\partial (nkT)}{\partial \vec{r}} = \frac{\partial \vec{p}}{\partial t}\Big|_{ibc} \tag{4.3.15}$$

$$\frac{\partial w}{\partial t} + \frac{\partial w}{\partial \vec{r}}.\vec{v} + \frac{1}{n}\frac{\partial (nkT\vec{v})}{\partial \vec{r}} + qn\vec{E}.\vec{v} - \frac{1}{n}\frac{\partial}{\partial \vec{r}}.(\kappa\frac{\partial T}{\partial \vec{r}}) = \frac{\partial w}{\partial t}\Big|_{ibc} \tag{4.3.16}$$

To account for the different relaxation times, momentum and energy relaxation times are introduced. The validity of the relaxation time approximation has been disscussed by Nougier. These approximations are basically valid in the presence of low energy scattering events. According to the relaxation time approximation:

$$\frac{\partial \vec{p}}{\partial t}\Big|_{ibc} = -\frac{\vec{p}}{\tau_p} \tag{4.3.17}$$

$$\frac{\partial w}{\partial t}\Big|_{ibc} = -\frac{w - w_0}{\tau_e} \tag{4.3.18}$$

These equations are suplemented by:

$$\vec{p} = m\vec{v} \tag{4.3.19}$$

$$w = \frac{3}{2}kT + \frac{mv^2}{2} \tag{4.3.20}$$

To proceed further, explicit values for energy and momentum relaxation times must be found. One can use the values of τ_e and τ_p as determined from Monte Carlo simulations. In any case, the values for the momentum and energy relaxation time, must satisfy the equations for energy and momentum, i.e. under steady state condition:

$$\tau_p(E) = \frac{m\vec{v}}{q\vec{E}} \tag{4.3.21}$$

$$\tau_e(E) = \frac{1}{\vec{E}.\vec{v}}[\frac{3}{2}k(T - T_o) + \frac{mv^2}{2}] \tag{4.3.22}$$

Most authors at this step, use these steady state values for the relaxation times, even under dynamic conditions. It is difficult to judge the validity of this assumption: The results predict the velocity overshoot, and it is very difficult to compare the numerical results to experimental values. Under steady state, $\vec{v} = \mu(E)\vec{E}$. For the mobility model, Baccarani and Woordeman use the empirical relation:

$$\mu(E) = \frac{\mu_0}{\sqrt{1 + (\mu_0 E/v_{sat})^2}} \tag{4.3.23}$$

To proceed further, the relation between carrier temperature and the applied electric field must be found. The Einstein relation is used:

$$T(E) = \frac{q}{k}\frac{D(E)}{\mu(E)} \tag{4.3.24}$$

The behavior of diffusivity vs. the electric field is relatively well known. It exibites slight anisotropy along the crystaline axis and small decrease at high electric fields. In most

simulation work this small variation in D is ignored, and the low field value is used:

$$D(E) = D_o = \frac{kT_o}{q}\mu_o \tag{4.3.25}$$

Finally for $T(E)$ one gets:

$$T(E) = T_o \frac{\mu_o}{\mu(E)} \tag{4.3.26}$$

where $\mu(E)$ was given before. Variations in effective mass in the electric field are ignored. Finally for the relaxation times in terms of carrier temperature, one gets:

$$\tau_p(E) = \frac{m}{q}\mu(E) = \frac{m\mu_0}{q}\frac{T_o}{T} \tag{4.3.27}$$

$$\tau_e(E) = \frac{m\mu_o}{2q}\frac{T_o}{T} + \frac{3}{2}\frac{k\mu_o}{qv_{SAT}^2}\frac{TT_o}{T+T_o} \tag{4.3.28}$$

4.3.2 Solution of the Coupled Equations

In what follows, a scheme for solution of the coupled equations will be developed. This scheme will be applied to a one dimensional system and compared to exact solution in one dimension. Both these solution techniques are different from the one in work by Baccarani and Wordeman.

From the first equation, for carrier conservation, one gets:

$$\nabla.(n\vec{v}) = 0 \tag{4.3.29}$$

This is the same as current continuity; $\vec{J} = -qn\vec{v}$, or $n\vec{v}$ =constant. Therefore in the equations for momentum and energy, either n or $\vec{v}$ can be eliminated, in terms of current.

The resulting equations are:

$$qD_o(1 - \frac{m|\vec{J}|^2}{kTq^2n^2})\nabla n + q\mu(\vec{E} + \frac{k}{q}\nabla T)n = \vec{J} \quad (4.3.30)$$

$$(\frac{5}{2} + r)\frac{k^2}{q}\nabla.(\mu nT\nabla T) + \frac{\vec{J}}{q}\nabla(\frac{5}{2}kT + \frac{mJ^2}{kTq^2n^2}) + \vec{E}.\vec{J} = \frac{3}{2}k\frac{n}{\tau_e}(T - T_o) + \frac{m|\vec{J}^2|}{kTq^2n\tau_e} \quad (4.3.31)$$

These are the momentum and energy conservation equations respectively. Equation (4.3.31) for the temperature is fairly easy to understand. Equation (4.3.30), is the same as the classical expression for current, but where an effective diffusion coefficient and electric field are introduced. The effective diffusion coeficient is:

$$D = D_o(1 - \frac{m|\vec{J}|^2}{kTq^2n^2}) \quad (4.3.32)$$

The extra term reduces the contribution of the diffusion term to the current (and velocity) in the presence of high carrier density gradients. To see this consider equation (4.3.30) in one dimension:

$$qD_o(1 - \frac{mJ^2}{kTq^2n^2})\frac{dn}{dx} + q\mu(E + \frac{k}{q}\frac{dT}{dx})n = J \quad (4.3.33)$$

From the Eq. (4.3.33) , the current density can be calculated:

$$J = \frac{q}{2\tau_p}\frac{n^2}{\frac{dn}{dx}} \times [\sqrt{1 + 4\tau_p(\mu E'\frac{1}{n}\frac{dn}{dx} + D_o\frac{1}{n^2}(\frac{dn}{dx})^2)} - 1] \quad (4.3.34)$$

where $E' = E + \frac{k}{q}\frac{dT}{dx}$. From the equation for current, the carrier velocity can be extracted as:

$$v = -\frac{1}{2\tau_p}(\frac{1}{n}\frac{dn}{dx})^{-1} \times [\sqrt{1 + 4\tau_p(\mu E'\frac{1}{n}\frac{dn}{dx} + D_o\frac{1}{n^2}(\frac{dn}{dx})^2)} - 1] \quad (4.3.35)$$

This is the expression for velocity, in the pressence of large gradients, $\frac{1}{n}\frac{dn}{dx}$. If $\frac{1}{n}\frac{dn}{dx}$ is small, a Taylor expansion of (4.3.35) can be made, and the classical expression for the carrier velocity is obtained:

$$v = -\mu(E)E - D(E)\frac{1}{n}\frac{dn}{dx} \tag{4.3.36}$$

Baccarani and Wordeman calculate carrier velocity vs $\frac{1}{n}\frac{dn}{dx}$. Their results are plotted for differnt values of E, in Fig. 4.1. The difference between the classical equation for velocity and the one obtained by including the non-stationary effects can be seen clearly. The classical equation, in the presence of large $\frac{1}{n}\frac{dn}{dx}$, places no bound on carrier velocity. Using the hydrodynamic equations, in the presence of large carrier density gradients is bounded by the thermal velocity:

$$v_{max} = \sqrt{D_o/\tau_p} = \sqrt{kT/m} = v_{Thermal} \tag{4.3.37}$$

which is expected from physical arguments. This is achieved by reduction in the effective diffusion coefficient in the presence of large carrier density gradients.

In one dimension Eq. (4.3.30) can be solved exactly. First the expression for current is obtained (Eq. 4.3.34)). Now one must solve:

$$\frac{\partial J}{\partial x} = 0 \tag{4.3.38}$$

It is difficult to linearize Eq. (4.3.34). Thus J is rewritten as:

$$J = \mu E(x)n(x) + D\frac{\partial n(x)}{\partial x} + F(n(x)) \tag{4.3.39}$$

where $F(n(x))$ is the correction to current due to large carrier density gradients:

$$F(n(x)) = \frac{q}{2\tau_p}\frac{n^2}{\frac{dn}{dx}} \times [\sqrt{1 + 4\tau_p(\mu E'\frac{1}{n}\frac{dn}{dx} + D_o\frac{1}{n^2}(\frac{dn}{dx})^2)} - 1] - \mu E(x)n(x) + D\frac{\partial n(x)}{\partial x} \tag{4.3.40}$$

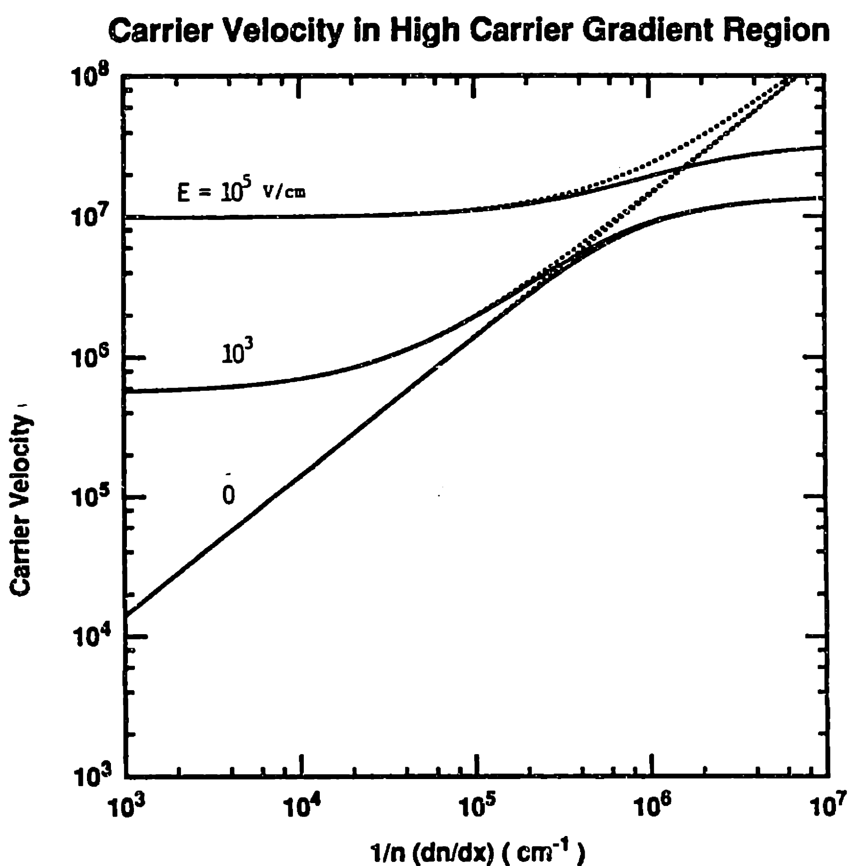


Figure 4.1: Carrier velocity in region of high carrier gradient as calculated by drift-diffusion and Hyrodynamic equations [Baccarani and Wordeman]. Solid line is full hydrodynamic equation and dashed line is obtained by DD

To solve equation (4.3.38) an initial guess is made on $n(x)$. This guess is used in equation (4.3.40) to calculate $\frac{\partial F}{\partial x}$ numerically. Then Eq. (4.3.38) is solved to obtain a new $n(x)$ and the cycle is repeated. This solution converges fairly rapidly (After 5 iteration, maximum change in $n(x)$ in each iteration is less than 10^{-3}.), and when coupled to the temperature equation, the solution to the complete set of hydrodynamic equations is obtained (The grid was uniform and the spacing between grid points aws 1 nm). This technique was applied to a step in electric field problem for an electron with initial mobility of 560 $cm^2.V/s$ (The conditions are similar to those in the work by Baccarani and Wordeman). Electrons enter a region of sharp discontinuity in electric field, i.e. $E = 10^3 V/cm$ for $x < 0.5\ \mu m$, and $E = 10^5 V/cm$ for $x > 0.5\ \mu m$ (Figure 4.2). Figure 4.3 presents the solution obtained using this technique (in solid curve). Equation for temperature, (4.3.31), is almost linear in T (τ_e has weak dependence on T). It was treated, and solved, as a linear equation by Gaussian elimination.

For a MOSFET structure the hydrodynamic equations, have to be coupled to the Poisson equation and be solved in two dimensions. The previous solution technique cannot be used in two dimensions. One simple way to solve (4.3.30) in two dimensions, is to introduce the effective diffusion coefficient according to Eq. (4.3.32) as discussed earlier, and then solve:

$$\nabla.\vec{J} = \nabla.[\mu\vec{E}(x,y)n(x,y) + D(x,y)\nabla.n(x,y)] = 0 \tag{4.3.41}$$

along with the Poisson equation in a standard way. Of course now D is a function of space, because of its dependence on $|J|$ and $n(x,y)$. This scheme (of using an effective D) was tested in one dimension and the result was compared to exact solutions. First an initial

guess on $n(x)$ was made. The current was calculated using:

$$J = q\mu En(x) + D_o \frac{\partial n}{\partial x} \tag{4.3.42}$$

Using $|J|$ and $n(x)$ a new $D(x)$ was calculated from Eq. (4.3.32). Using the new D, the following equation:

$$\nabla.\vec{J} = \nabla.[\mu\vec{E}'(x)n(x) + D(x)\nabla.n(x)] = 0 \tag{4.3.43}$$

was solved in one dimension. The procedure was iterated until the solution for n and T converged. As before a grid size of 1 nm was used, andsame technique was applied to solve for temperature. It took about 5 iterations as before. This technique was applied to the step in electric field problem of Fig. 4.3. For the porpose of comparision, the final v was calculated exactly using Eq. (4.4.45). The result is superimposed in Fig. 4.4 on top of that obtained by exact solution in dashed line (It is hard to distinguish them).

These equations were also applied to the same step problem, but the initial mobilities were varied. As can be seen in Fig. 4.4, as the initial mobility is increased, the carrier velocity increases, and the velocity overshoot becomes more pronounced.

In summary, the hydrodynamic model of transport was presented here. It is based on the first and second moments of BE. Although, it is based on a Maxwellian distribution assumption, it is capable of predicting some of the non-stationary effects, and it provides a simple way of accounting for the effects caused by the rapidly varying fields.

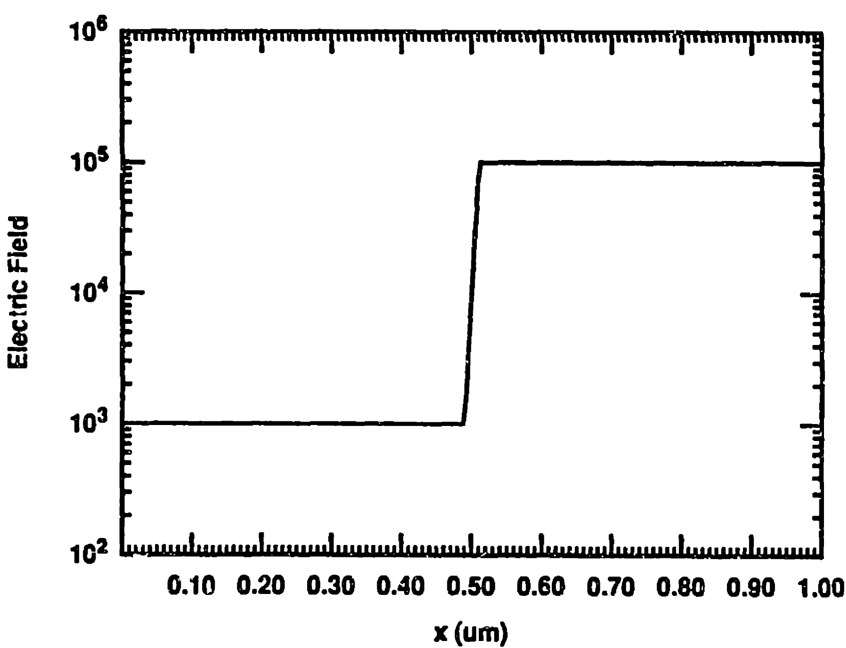


Fig. 4.2 Elecctric field profile used to study non-stationary effect caused by rapid variations in electric field

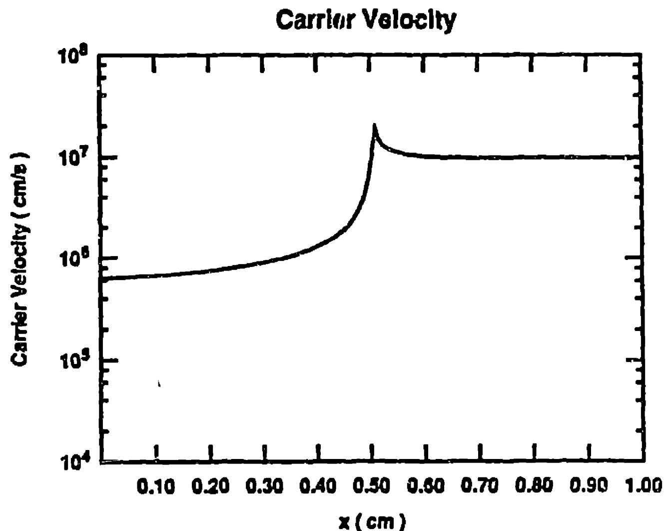


Fig. 4.3a. Carrier velocity in a region of sharp electric field discontinuity (Solid line is exact and dashed line is using effective D)

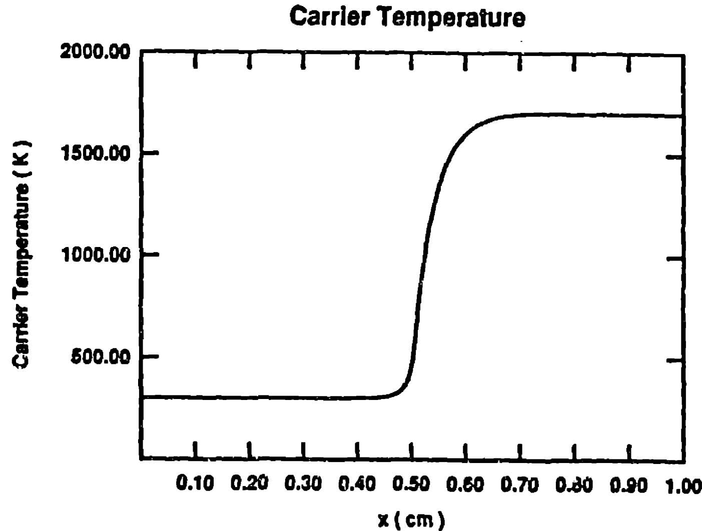


Fig.4.3b Carrier temperature in a region of sharp electric field discontinuity (Solid line is exact and dashed line is using effective D)

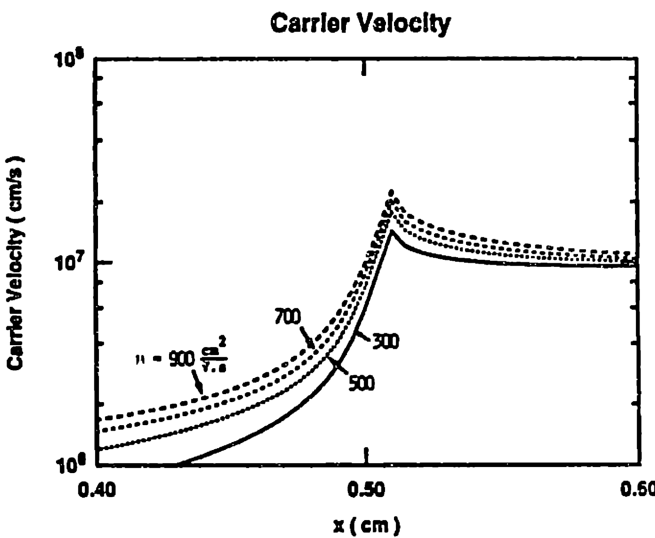


Fig. 4.4a. Carrier velocity in a region of sharp electric field discontinuity for different mobilities

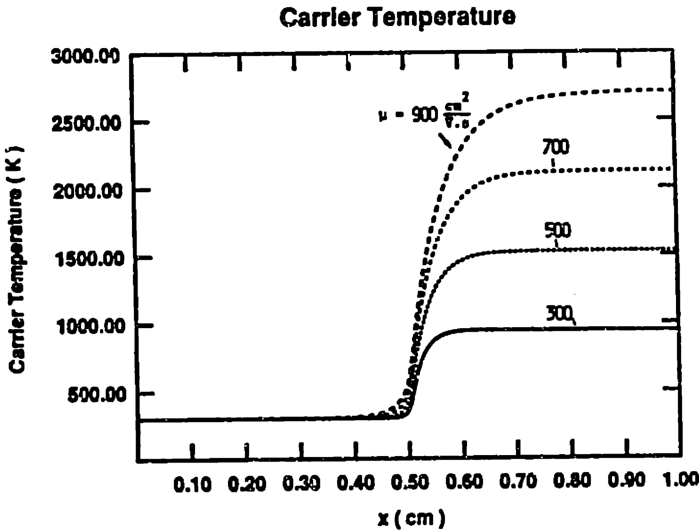


Fig.4.4b Carrier temperature in a region of sharp electric field discontinuity field discontinuity for different mobilities

Chapter 5

Velocity Overshoot

Carrier transport in high field regions is of great theoretical and practical interest. One of the basic measured quantities is carrier velocity. By relating the velocity to the electric field, information about carrier dynamics can be obtained. Under steady state:

$$v_e = \mu E + D \frac{1}{n} \frac{\partial n}{\partial x}$$

where $\mu = \mu(E)$ and $D = D(E)$, i.e. the carrier mobility and diffusitivity is a *local* function of electric field. In the high field region, and during short distances (or times), as we saw previously, the carrier mobility is a function of the carrier energy (not of the local field), and as the carrier energy changes because of the presence of the high field (during energy relaxation time, τ_e) so does the mobility (within momentum relaxation time, τ_p). One result of this is the phenomenon of velocity overshoot. The importance of observation of velocity overshoot does not only lie in the confirmation of an slight increase in carrier velocity, but also that indeed:

$$\mu = \mu(T(E))$$

This chapter presents the results that indicate velocity overshoot takes place at room temperature and at 77 K.

5.1 Velocity Overshoot

When electrons enter a region of high field in Si, their velocity can exceed the steady state saturation value for a short period of time, τ_e, the energy relaxation time. This phenomenon has been predicted theoretically [Rauch], and has been observed experimentally at 4.2 K by Chou et al [Chou, 1985], based on measurements of the saturated transconductance of sub-100 nm channel-length Si MOSFET's.

Under stationary transport conditions the electron drift velocity can be written as $v_e = \mu_e E$, where μ is the mobility. Mobility decreases as electric field is increased, with the result that the carrier velocity saturates. Assuming a Maxwellian electron distribution in phase space, μ can be written as [Sttraton]:

$$\mu_e(T_e) = \mu(T_\circ) \times [\frac{T_\circ}{T_e}]^\alpha \tag{5.1.1}$$

where T_e is defined as a measure of the spread of the Maxwellian distribution, $T_\circ$ is the lattice temperature, $\mu(T_\circ)$ is the low field mobility, and α is an exponential factor that determines the mobility behavior as a fuunction of T_e. When electrons move from a low field region into a high field region, their temperature increases but does not reach its final value instantaneously. Instead it takes about the energy relaxation time, τ_e, during which the mobility is also higher than its final steady state value. Therefore,

for a time comparable to τ_e , one can have high mobility electrons in a region of high electric field. During this time the carrier velocity can exceed the final steady state value. The higher the initial mobility, the higher this velocity overshoot will be.

In addition to high mobility carriers, a region of high and abrupt electric field is needed for the velocity overshoot to occure. MOSFET is the ideal structure for both measurement of carrier velocity and observation of velocity overshoot. In MOSFET's, the cold carriers are injected from the source in to the high field region of the channel. For short channel MOSFETs, the drain current is usually written as [Sodini]:

$$I_D = W v_{eff} C_{ox}(V_{GS} - V_T)$$

where v_{eff} is the effective carrier velocity in the channel. It has to be emphasized that v_{eff} is the *minimum* carrier ensemble velocity in the channel, since the maximum gate drive, anywhere in the channel, is $C_{ox}(V_{GS} - V_T)$. This maximum drive can only be obtained if all the charges are imaged on the gate (and this is usually not the case). Furthermore it has to be recognized that maximum drive occures at the source, where the surface potential is the lowest. Thus v_{eff} can also be viewed as the velocity at the source. Thus in order to be able to observe velocity overshoot, in addition to high carrier mobility, the lateral field at the source must be increased.

To increase the lateral field at the source, the channel length must be decreased. This is shown in Fig 5.1 and 5.2. In Fig. 5.1, the lateral field in a 0.5 μm device and its corresponding carrier velocity is presented. As it can be seen, since the lateral field at the source is not high enough, velocity at the source is way below its bulk

saturation value. Fig. 5.2a, where the lateral fields for a 0.1 μm and 0.15 μm device for V_{DS} = 2.0 V are plotted (obtained using MINIMOS). It can be seen that as L is reduced, the lateral field at the source increases.

High field at the source results in high velocity at the source. This is shown in Fig. 5.2b, where the carrier velocities corresponding to Fig. 5.2a is shown. Velocities were calculated by applying one-dimensional hydrodynamic equation to fields obtained at the Si/SiO_2 interface from MINIMOS simulation. Carrier mobility was assumed to be 500 $cm^2/V.s$. Velocity still is below (but very close) the bulk saturation value for a 0.1 μm device. If the mobility is increased to 700, velocity overshoot is observed at the source. As it will be shown, velocity overshoot is observed using devices with mobility sligthly below 500. The inability of this very simple procedure to show velocity overshoot is probably due to bad modeling of relaxation times. It has to be emphasized that once the device enter saturation, any increase in V_{DS} will affect the drain region of the device, only.

To measure the velocity in device, its intrinsic transconductance must be measured. In short-channel MOSFET's, the intrinsic transconductance, g_{mi}, can be related to the average carrier velocity v_e, by $g_{mi} = C_{ox}v_e$ where C_{ox} is the oxide capacitance.

5.2 Experimental Result

Based on classical scaling theory, short channel MOSFET's have previously been fabricated on relatively heavily doped substrates. Heavy doping increases the normal field at which inversion occurs, and thus reduces the channel mobility [Sun]. As doping

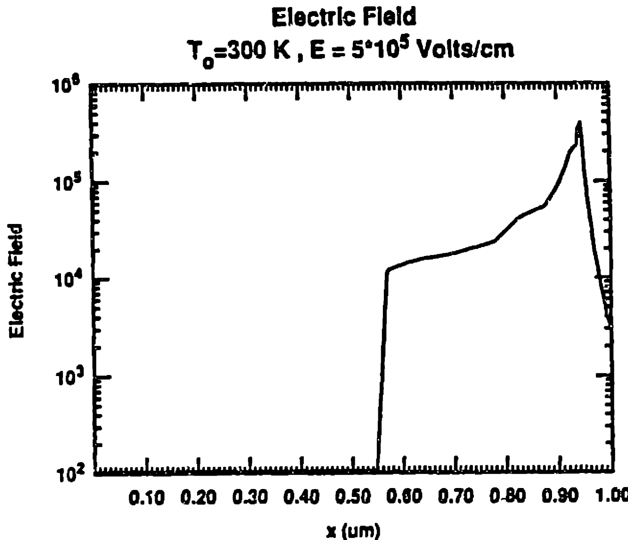


Figure 5.1a: Lateral field for a 0.5 μm device for $V_{DS} = 2.0$ V

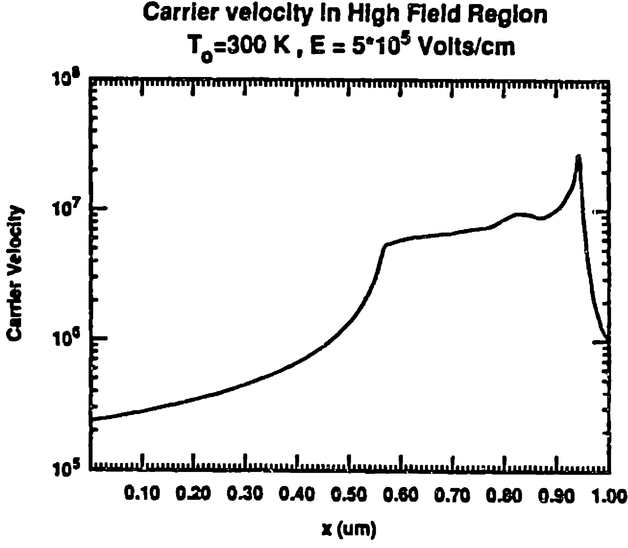


Figure 5.1b: Carrier velocity for a 0.5 μm device for $V_{DS} = 2.0$ V

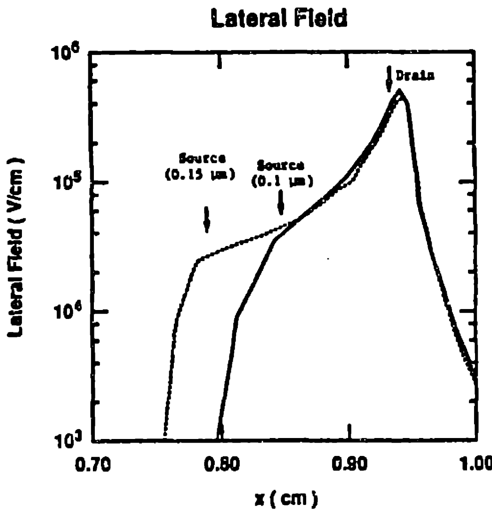


Figure 5.2a: Lateral fields for a 0.1 μm and 0.15 μm device for $V_{DS} = 2.0$ V

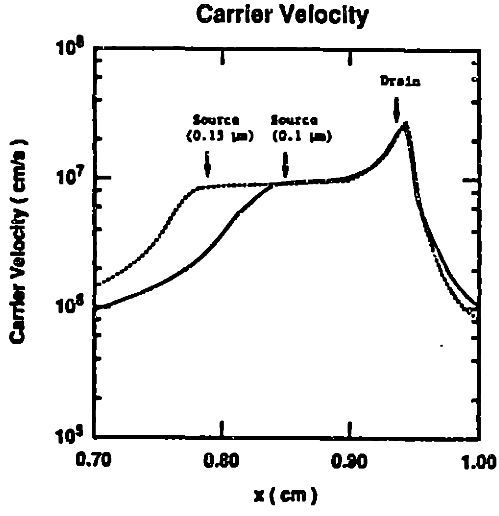


Figure 5.2b: Carrier velocities for a 0.1 μm and 0.15 μm device for $V_{DS} = 2.0$ V

increases above a certain level, increased impurity scattering also contributes to reduction of mobility. To increase the mobility, a non-uniformly doped channel was used in this work. This nonuniform doping permits high surface mobility to be achieved while avoiding punchthrough. To achieve the non-uniform doping, a boron channel implant of 5×10^{12} cm^{-2} at 50 KeV and a 10 min anneal at 900°C was used. This gave a surface doping of 2×10^{16} cm^{-3}. The boron concentration rises with distance from the surface, reaching a peak of 2×10^{17} cm^{-3} at 0.17 μm depth. The profile of boron is presented in Fig. 5.3a. The remainder of the fabrication procedure, which utilizes an X-ray lithography step, was similar to that reported in Chapter 3. The gate oxide was 7.5 nm thick, drain junction depth was estimated to be about 0.11 μm. The fabricated devices had well-behaved characteristics down to channel lengths of 1000 μm. Channel lengths and source drain resistivity were electrically measured as outlined by [Suciu and Johnston]. To verify the extracted channel length, a two dimensional device simulator (MINIMOS) was used to model the main short channel effect, the threshold roll-off at short channel lenghts. Figure 5.3b shows the behavior of the threshold voltage, V_T, vs. channel length for non-uniform channel implant. Both experimental results and results from a two-dimentional device simulator ,MINIMOS [Selberher], are plotted. With the exception of a uniform V_T difference of 100 mV, the simulation matches the experimental data quite well; in particular, the onset of V_T roll-off with channel length is in good agreement. Also shown in Fig. 5.3b are V_T calculations for a uniform channel doping of 2×10^{17} cm^{-3}. The non-uniform doping

did not cause any increase in short channel effects; yet, because of the reduction in surface doping we obtained a low field mobility of 450 $cm^2/V.s$, versus 300 $cm^2/V.s$ that was measured for uniform doping [Chou].

The measured source drain parasitic resistance was also measured electrically. R_{SD} was 1400 $\Omega - \mu m$ at room temperature and 1000 $\Omega - \mu m$ at 77 K. The measured value was independent of the measured channel length , and was fairly uniform across the wafer.

5.2.2 Observation of Overshoot

Typical device characteristics are shown in Fig. 5.4 for a device with W=8 μm and L=90 nm, at 300 K and 77 K. The device has clear saturation and non-saturation regions of operation. The transconductance of the device increases by a factor of 1.5 between 300 K and 77 K. It is noteworthy that the bulk saturation velocity increases by only 1.3 between 300 K and 77 K. Therefore if the velocity is just saturated at 300 K, the velocity at 77 K must represent a velocity overshoot.

To obtain the average carrier velocity, we calculated the intrinsic tranconductance of the devices ,g_{mi}, from the measured tranconductance g_m and the measured source-drain resistance R_{SD} [Chou, 1987]:

$$g_{mi} = g_{mo}/(1 - R_{SD} g_d (1 + R_S g_{mo})) \tag{5.5.1}$$

where $g_{mo} = g_m/(1 - R_S g_{mo})$, R_S is the source resistance, and g_d is the output conductance at $V_{DS} = 1.5\ V$. Figure 5.5(a) shows the measured transconductances of

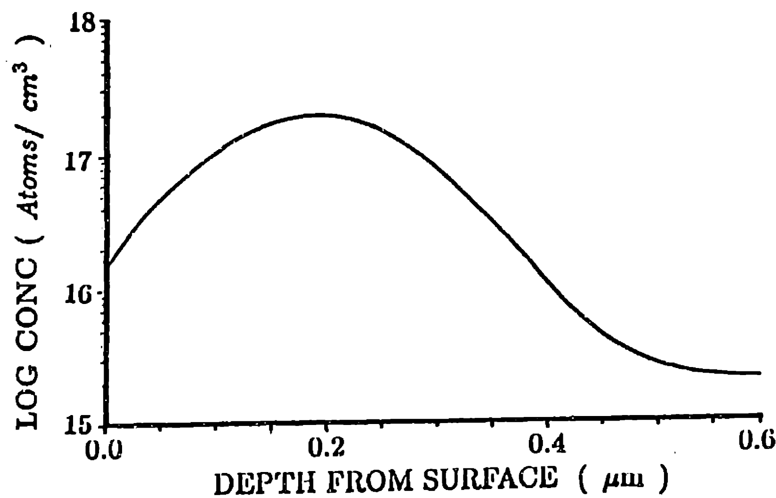


Figure 5.3a : Channel doping used in observation of velocity overshoot

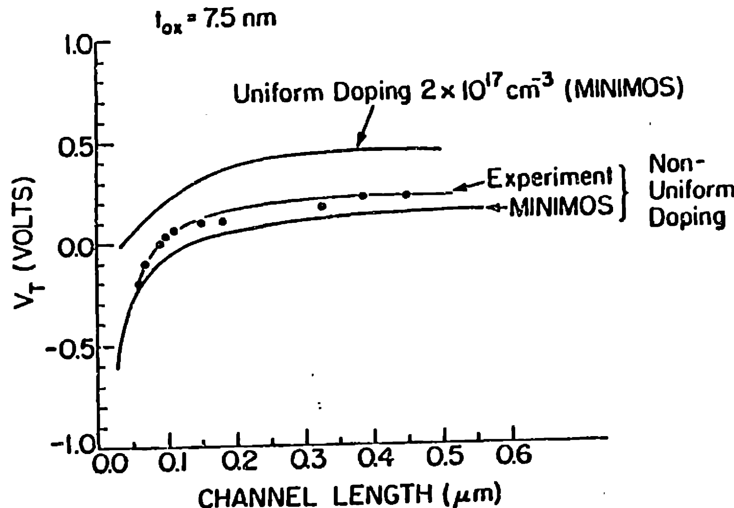


Figure 5.3b : Threshold voltage vs. channel length for uniform and nonuniform substrate doping, as simulated by a two dimensional device simulator. The surface doping for the nonuniform case is 2×10^{16} cm^{-3}. Experimental results for the nonuniform doping case are also shown.

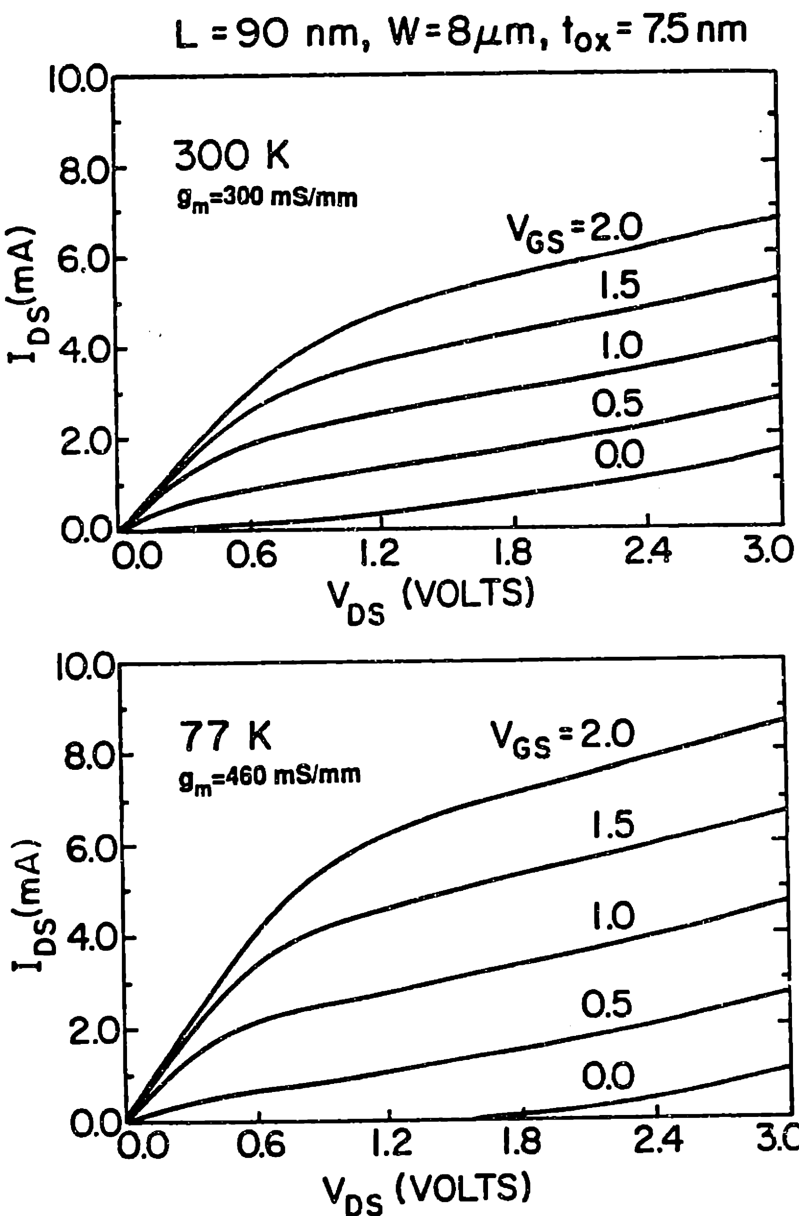
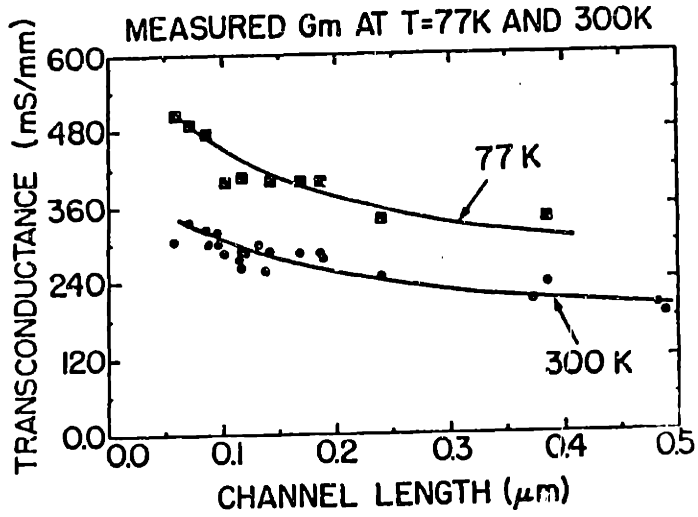


FIG. 5.5 : Charactristics of a 90-nm-channel length device at (a)300 K and (b) 77 K.

the devices at 77K and 300 K. The calculated average carrier velocity was obtained from the intrinsic transconductance using the relation $v_e = g_{mi}/C_{ox}$. The results are plotted in Fig. 5.5(b). As can be seen, for channel lengths of less than 1000 μm at 300 K, and less than 1500 μm at 77 K, the average channel velocity exceeds the bulk saturation value. Chou *et al* have reported a maximum v_e of 0.75×10^7 cm/s at room temperature with a low field mobility of 300 $cm^2/v.s$, while a maximum v_e of 1.3×10^7 cm/s was obtained with μ_e of 450 $cm^2/V.s$.

In conclusion, by using non-uinform channel doping, high mobility short channel MOSFETs were fabricated, and using these devices velocity overshoot for the first time was observed at 77 K and room temperature. This is confirmation of a major non-stationary effect which was predicted over a long time.



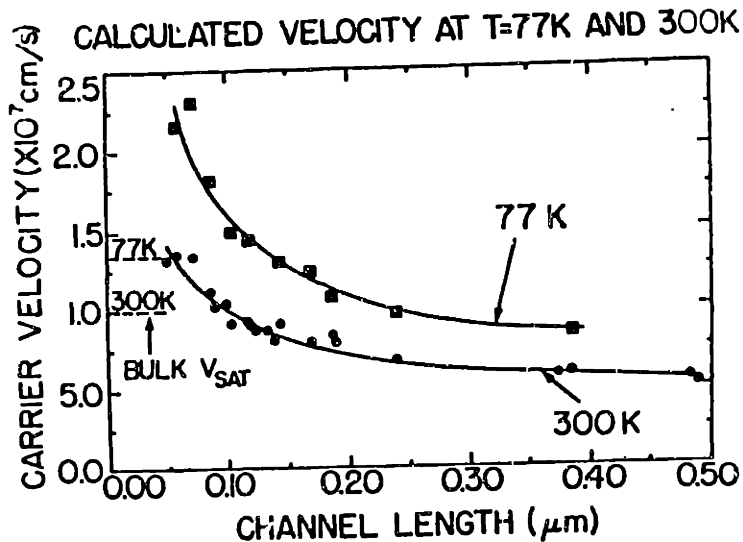


FIG. 5.5 : (a) Measured transconductances of devices vs channel lengths at 300 K and 77 K. (b) Calculated average velocities at 300 K and 77 K.

Chapter 6

Hot Electron Effects

6.1 INTRODUCTION

As MOS transisitor dimensions are scaled down, the devices operate under higher electric fields and current densities. High fields and currents are responsible for hot carrier effects such as substrate current, gate current, and light emission. Channel hot carriers impacting the Si/SiO_2 interface lead to degradation of the transconductance of the device, threshold shift, shift in the subthreshold current slope, source-drain breakdown, and finit device lifetime [Hu]. These effects have been studied extensively for devices with channel lengths, L $\geq$ 0.5 μm. In this work, the hot electron generated substrate current in deep submicron devices is studied. The observation of velocity overshoot, and understanding of that phenomenon through the reduction in the carrier temperature was the motivation behind the investigation of the hot electron effects in deep submicron devices. Specifically, the goal was to look at the hot electron effect in the channel length region where the velocity overshoot was observed and transport was strongly affected by the non-stationary effects. Substrate current generation was used as an indicator of the amount of hot electron generation: It is easy to measure substarte current and it is well understood. This chapter first presents the lucky electron model of impact ionization [Shockley]. Then a quasi two-

dimensional model of the lateral channel field is presented. Using the lucky electron model and the lateral field model, the lucky electron model of substrate generation is presented. Next data will be shown that support the model down to L=0.5 μm. The measurements in devices with L<0.5 μm are presented which show effects unique to short channel devices.

6.2 Avalanche Multipication

High electric fields are the primary cause for hot electron effects in a variety of devices. Carriers gain energy as they trevlel in their drift field, and because of the scattering events, their steady state energy and temperature saturates and is low. They do not have the energy necessary to cause a hot electron effect event. If the field is high enough, elctrons can gain enormous amount of energy in a short distance (of the same order of their scattering length), and some of them (the lucky ones) do not go through scattering events. They gain enough energy to surpass some threshold necessary for a hot electron event. To model hot electron effects, one must calculate the probability that an electron would travel a distance to gain enough energy (threshold energy for the hot electron process), without loss of that energy through some scattering mechanism.

One of the most common hot electron effects is impact ionization, i.e. generation of electron-hole pairs in high fields. This happens when carriers gain enough energy (1.1 eV in Si) to send an electron from the valance to the conduction band. The basic model used is due to Shockley. Given:

E_t: The threshold energy measured from band edge for a carrier above which it may produce a hole-electron pair.

λ: The mean free path for optical phonon scattering. Optical phonon emission is the primary loss mechnism at high energies.

L_i: The mean free path between ionization for carriers with $W > E_i$.

then the impact ionization coefficient $\alpha(E)$, (defined as pairs produced per cm in an electric field of E) is given by:

$$\alpha(E) = (\frac{qEL_i}{\lambda E_R})exp(-E_i/q\lambda E) \tag{6.2.1}$$

This expression is used to calculate the amount of substarte current in Si MOSFET according top the expression :

$$I_{SUB} = \int_{channel} \alpha(x,y) I_{DS} dxdy \tag{6.2.2}$$

To calculate the integral, the profile of the electric field in the channel muet be known.

6.3 Lateral Electric Field

As carriers travel from source to drain, they gain energy $\int \vec{E}.\vec{dx}$, and the carrier energy is increased. The impact ionization anywhere in the channel is locally proportional to the energy. To calculate the amount of the impact ionization, it is important to know the shap of the elctric field in the channel.

When the drain to source bias, V_{DS} is low (in the linear region), the inversion layer extends all the way to the drain, and the lateral electric field is low. As V_{DS} is increased, for $V_{DS} > (V_{GS} - V_T)$, at the drain edge, the normal field reverses sign and pushes the carrier toward the bulk (pinch-off point). At that point the lateral electric field also increases dramatically and causes velocity saturation (saturation point). Thus carriers begin to move away from the surface, into the bulk. This phenomenon gives rise to current saturation. (It should be noted that the pinch-off and saturation points may not coincident, but in the analysis that follows, it is assumed that they are.) In a device in saturation, the lateral electric field grows exponentially from the saturation point to the drain edge.

It will be shown that the exponential rise is caused by the the carriers being pushed into the bulk or field reversal at the pinch-off point (or divergence normal field). Thus finding the electric field profile, is inherently a two dimensional problem.

To start, Poission's equation along with current continuity equation must be solved:

$$\nabla \cdot \vec{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = \frac{-(qN_{SUB} + Q_I)}{\epsilon_{Si}} \tag{6.3.1}$$

where Q_{MOB} is the mobile charge density, y is the direction along the channel, and x is the direction normal to the channel. A simple approximation is to ignore $\frac{\partial E_x}{\partial x}$ and just solve for the field across a reverse-biased pn junction, and include a fraction of the mobile charge. This is the so called constant field gradiednt approximation:

$$\frac{\partial E_y}{\partial y} = \frac{-(qN_{SUB})}{\epsilon_{Si}} + \frac{\alpha Q_{MOB}}{\epsilon_{Si}} \tag{6.3.2}$$

where α is the fraction of the mobile charge that contibutes to the lateral field. This equation results in a linear lateral field, which is very different from what is obtained by the two dimensional simulator.

There is a pseudo-two dimensional analysis due to Elmansy and later modified by Ko which gives a very elegant and simple (yet remarkably accurate) method of finding the maximum lateral field in the channel. In this method, the lateral field in the saturation region of a device (operated in saturation) is found.

6.3.3 Derivation of lateral Field

Derivation of lateral field, due to Al-Mansy is presented in Appendix A. It is shown that the mobile charge has makes a small contribution to the lateral field, and the most important contribution is from the divergence of the normal field. P.K. Ko used a slighly simplified version of El-Mansy's technique and obtained the lateral field in the channel. In

his calculations, he included the effect of finite junction depth in a heuristic way. His results are widely used. Consider the device whose saturation is shown in Fig. 5.2a. V_{DSAT} id defined as the saturation voltage, i.e. the voltage after which all the carriers move with the saturation velocity. Point C' is the lowest point to which the mobile charges are pushed by the reverse gate bias at the drain edge. B'C' is the locus of the electric field originating fron the B' point and ending on the space charge layer. Thus all the charges inside the AB'C'D are controled by the field lines crossing the 3 surfaces of AD, DC' and AB'. Let $V_s(y)$ be the surface potential at distance y from A and $E_s(y)$ be the corresponding channel field. Writting the Gauss's law on the surfce shown and going through the same procedure as before, one obtaines:

$$\frac{\partial}{\partial y}\int_0^{x'(y)} E_x(x,y)dx + \frac{\epsilon_{ox}}{\epsilon_{Si}}E_{ox}(o,y) =$$
$$\frac{q}{\epsilon_{Si}}[N_{SUB}\frac{\partial}{\partial y}\int_0^{y} x'(y")dy"] + \frac{Q_{MOB}}{\epsilon_{Si}}$$

This is the same equation as in Appendix A, but with the difference that the limit of the integral is the spread of the lateral field at y, $x'(y)$, instead of W_{depl}. Since $x'(y)$, is not known, the above equation can not be solved. Ko and Elmansy take different approaches to proceed. Elmansy assumes that x' is constant and is given by $\frac{x_j}{3}$, where x_j is the space charge layer thickness under the gate. Ko also uses a rectangular approximation to the gaussian surface, and assumes that $x'(y) = (x_j + \Delta x_j)$ where Δx_j is a depth modification due to junction depth and substrate doping. To first order Ko assumes that Δx_j depends on the difference between x_1 and x_j. Also Δx_j is supposed to aproach zero in the limiting case that AB' boundary appproaches the drain junction. Thus Ko uses the following approximation:

$$\Delta x_j = K_j(V_{DS} - V_{DSAT})(x_1 - x_j) \tag{6.3.3}$$

and K_j is a proportionality constant to be determined. Then the resulting equations

become:

$$\frac{\partial}{\partial y}\int_0^{(x_j+\Delta x_j)} E_x(x,y)dx + \frac{\epsilon_{ox}}{\epsilon_{Si}}E_{ox}(o,y) =$$
$$\frac{q}{\epsilon_{Si}}[N_{SUB}(x_j+\Delta x_j)] + \frac{Q_{MOB}}{\epsilon_{Si}} \tag{6.3.4}$$

which is the same as before. The integral $\int_0^{(x_j+\Delta x_j)} E_x(x,y)dx$ depends on how fast the lateral field spreads into the bulk. Elmansy assumes that (6.3.12):

$$\int_0^{(x_j+\Delta x_j)} E_x(x,y)dx = \frac{x_j}{3}E_s \tag{6.3.5}$$

Ko uses a more general formula:

$$\int_0^{(x_j+\Delta x_j)} E_x(x,y)dx = \frac{x_j}{n_{sp}}E_s \tag{6.3.6}$$

where the larger n_{sp}, the faster the field decayes into the bulk. Finally for the diffrential equation for the lateral field Ko gets:

$$\frac{(x_j+\Delta x_j)}{n_{sp}}\frac{\partial E_s(y)}{\partial y} + \frac{\epsilon_{ox}}{\epsilon_{Si}}\frac{(V_{GS}-V_T-V_s(y))}{t_{ox}} =$$
$$\frac{q}{\epsilon_{Si}}[N_{SUB}(x_j+\Delta x_j)] + \frac{Q_{MOB}}{\epsilon_{Si}} \tag{6.3.7}$$

where one uses $E_s(y) = \frac{(V_{GS}-V_T-V_s(y))}{t_{ox}}$

To proceed further Ko ignores all the mobile and imobile charges (similair th El-Mansy) and for the final differential equation he gets:

$$\frac{\partial E_s(y)}{\partial y} = A^2(V_S(y) - V_{DSAT}) \tag{6.3.8}$$

where

$$A^2 = \frac{\epsilon_{ox}n_{sp}}{\epsilon_{Si}t_{ox}(x_j+\Delta x_j)} \tag{6.3.9}$$

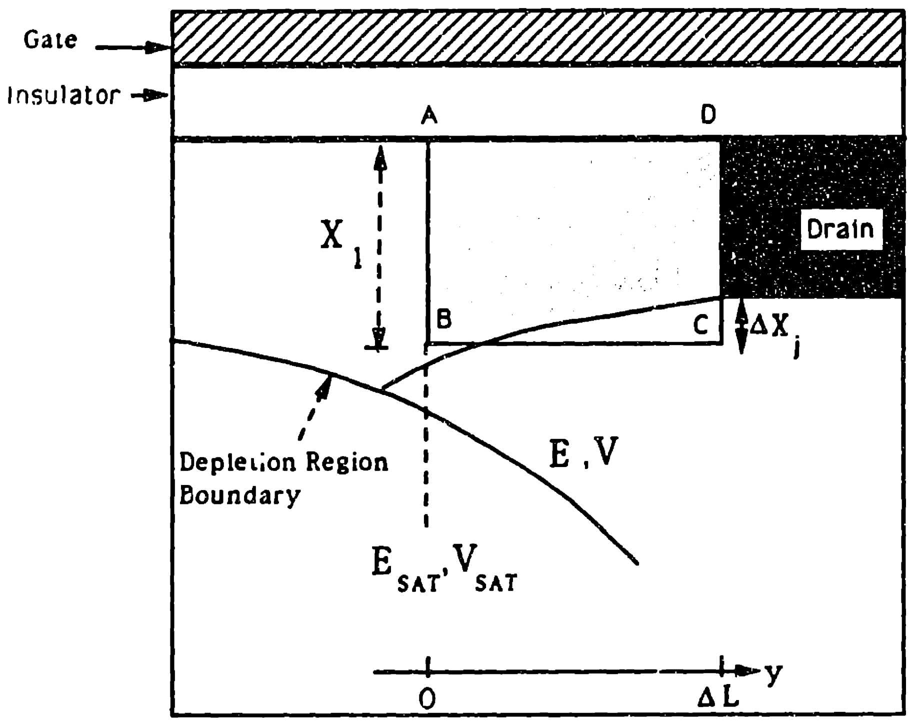


Figure 6.1: Drain region of A MOSFET as used by Ko

[To get the final differential equation in his work, Ko uses the wrong equation: $\frac{(V_{GS}-V_T-V_s(y))}{t_{ox}} = \frac{1}{C_{ox}}[N_{SUB}(x_j+\Delta x_j)+Q_{MOB}]$. For a device in saturation, such relation no longer holds.] It was shown previously that ignoring the mobile charges is valid when operating the device in the strong V_{DS} region and for long channel lengths. Solution to this equation is:

$$E_s(y) = E_{SAT}\cosh(Ay) \tag{6.3.10}$$

Imposing the boundary conditions at point A ($E_S = E_{SAT}$ for $y = 0$) and point B at the drain ($E_S = E_D$ for $y = \Delta L$), Ko gets:

$$(V_s(y) - V_{DSAT}) = (\frac{E_{SAT}}{A})\sinh(Ay) \tag{6.3.11}$$

and field at the surface becomes:

$$E_s = \sqrt{A^2(V_s(y) - V_{DSAT})^2 - E_{SAT}^2} \tag{6.3.12}$$

Both the lateral field and the potential in the channel are known, except for some constants. Ko assumes that the effect of substrate doping is small: $K_j = 0.01$. n_{sp} is estimated to be around 1.0-1.2, using a two dimentional simulator.

6.4 Substrate Current

When carriers go through the high field region of the channel, some of them acquire enough energy to cause impact ionization. Their fraction is given by the impact ionization coefficient. The total substrate current is given by:

$$I_{SUB} = \int_0^L I(x,y)\alpha(x,y)dxdy \tag{6.4.1}$$

where $\alpha(x,y) = \alpha(E(x,y))$ is the impact ionization coefficient. For the impact ionization coefficient, many authours use the following form:

$$\alpha(E) = \alpha_o e^{-\frac{\beta}{E}} \tag{6.4.2}$$

Strictly speaking, this is different from the form obtained by Shockley. Many measurement show an almost exponential dependence on $\frac{1}{E}$. Assuming that all the carriers go through the same high field region, the expression for I_{SUB} can be rewritten as:

$$I_{SUB} = I_D \int_0^L \alpha_0 \exp(-\frac{\beta}{E_S(y)})dy \qquad (6.4.3)$$

Using $E_s = \sqrt{4^2(V_s(y) - V_{DSAT})^2 - E_{SAT}^2}$, the substrate current becomes:

$$I_{SUB} = \frac{I_D\alpha_o}{A}\frac{E_D}{\beta}e^{-\frac{\beta}{E_D}} \qquad (6.4.5)$$

According to this expression, substrate current only depends on the lateral field at the drain and the drain current. This model can be further simplified by noticing that E_D depends on the difference between the drain-to-source voltage and the drain saturation voltage, $(V_{DS} - V_{DSAT})$, according to

$$E_D = \sqrt{A^2(V_{DS} - V_{DSAT})^2 + E_{SAT}^2} \qquad (6.4.6)$$

For large drain to source bias, an expansion of the following relation can be made:

$$E_D = (V_{DS} - V_{DSAT})/l \qquad (6.4.7)$$

where $l = \frac{1}{A}$ is a length parameter inversely proportional to the gradient of the lateral field at the drain edge. For long channel devices (from (6.3.9)):

$$l \approx (3t_{ox}X_j)^{-0.5} \qquad (6.4.8)$$

where t_{ox} is the oxide thickness, and X_j is the junction depth. The substrate current then becomes:

$$I_{SUB} = \frac{I_D\alpha_o}{\beta}(V_{DS} - V_{DSAT})e^{-\frac{\beta}{A(V_{DS}-V_{DSAT})}} \qquad (6.4.9)$$

This expression has worked remarkably well over a very wide range of channel lengths, and provides a unified method of characterizing the substrate current.

6.5 Experimental Results at 300 K

Hot-electron-generated substrate currents were measured in MOSFETs with effective channel length, L, ranging from 0.06 μm to 5 μm. These devices had a non-uniformly doped channel (Boron, 60 KeV, Dose=5×10^{12}). The oxide thickness was 4.8 nm. Total parasitic source-drain resistance was $R_{SD} = 850\ \Omega\mu m$ at room temperature and 650 $\Omega\mu m$ at 77 K. There was 10% variation in R_{SD} across wafer. Channel lengths were measured electrically, and they correlated closely with the drawn PMMA resist line width that defined them.

Fig. 6.2 demonstrates the essential observation: In Fig. 6.2(a) the drain current, I_D, and the substrate current,I_{SUB}, for two long channel devices are presented. As expected, as L is reduced both I_D and I_{SUB} increase. The ratio of substrate current to drain current, I_{SUB}/I_D, remains essentially constant as a function of V_{DS}. But for very small channel lengths, as shown in Fig. 6.2(b), as L is reduced from 0.14 μm to 0.12 μm, although I_D (and the transconductance) increases, I_{SUB} and therefore I_{SUB}/I_D decrease.

Note the increase in both drain and substrate current in the long channel regime as L is reduced. In the short channel regime, I_{SUB} decreases, even though as L is reduced I_D increases.

To present the observations in a more systematic way, the lucky electron model was used, and I_{SUB}/I_D was plotted vs. $(V_{DS} - V_{DSAT})^{-1}$, as shown in Fig. 6.3 (V_{DSAT} was determined using the method described in [Ko, 1980]). At a given $(V_{DS} - V_{DSAT})^{-1}$, for $0.5\mu m < L < 5\mu m$, I_{SUB}/I_D remains the same, irrespective of the gate bias (measured). As L is reduced below $0.5\mu m$, I_{SUB}/I_D increases. However, for $L < 0.15\mu m$, I_{SUB}/I_D decreases.

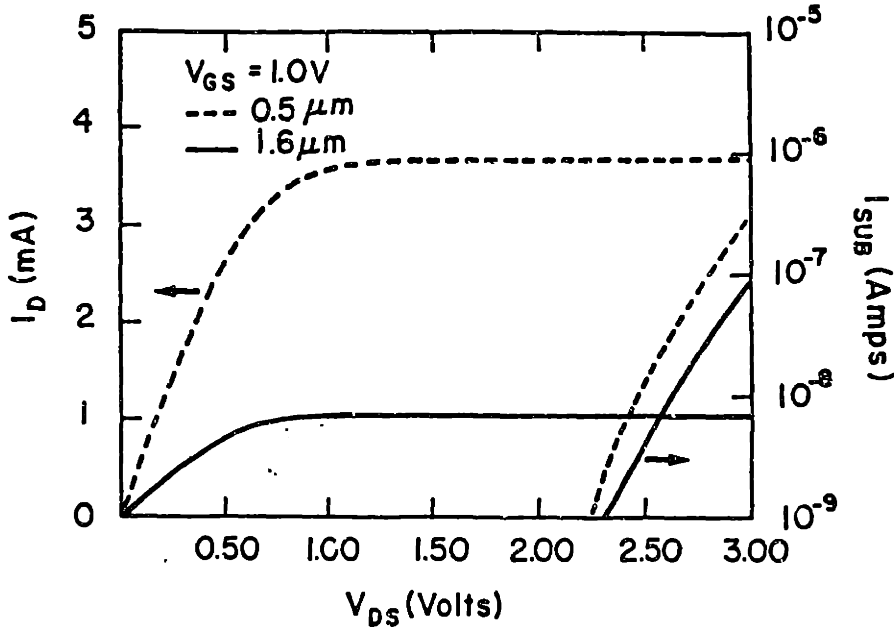


Figure 6.2a: Drain and substrate current at room temperature for two long channel MOSFETs.

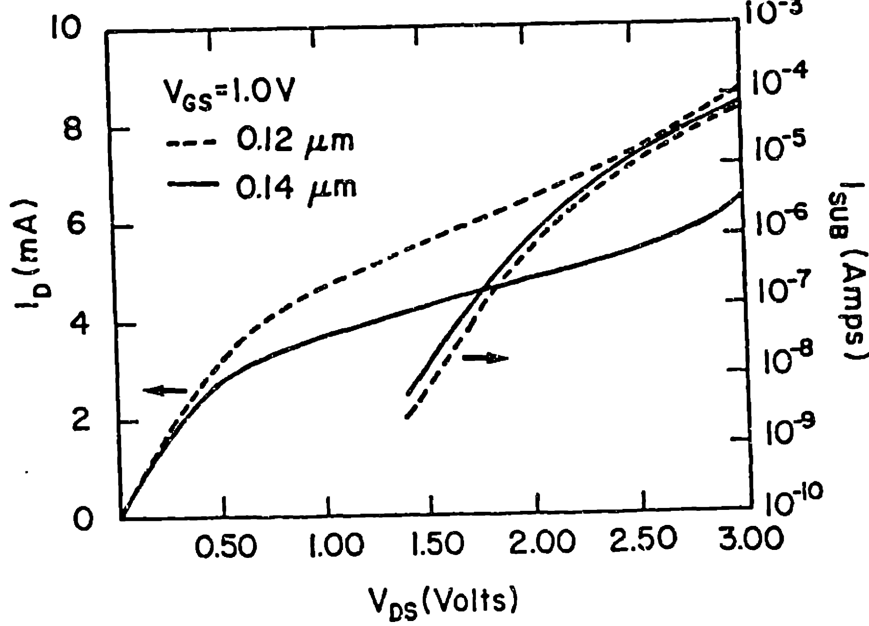


Figure 6.2b: Drain and substrate current at room temperature for two two short channel MOSFETs under identical bias as in (a)

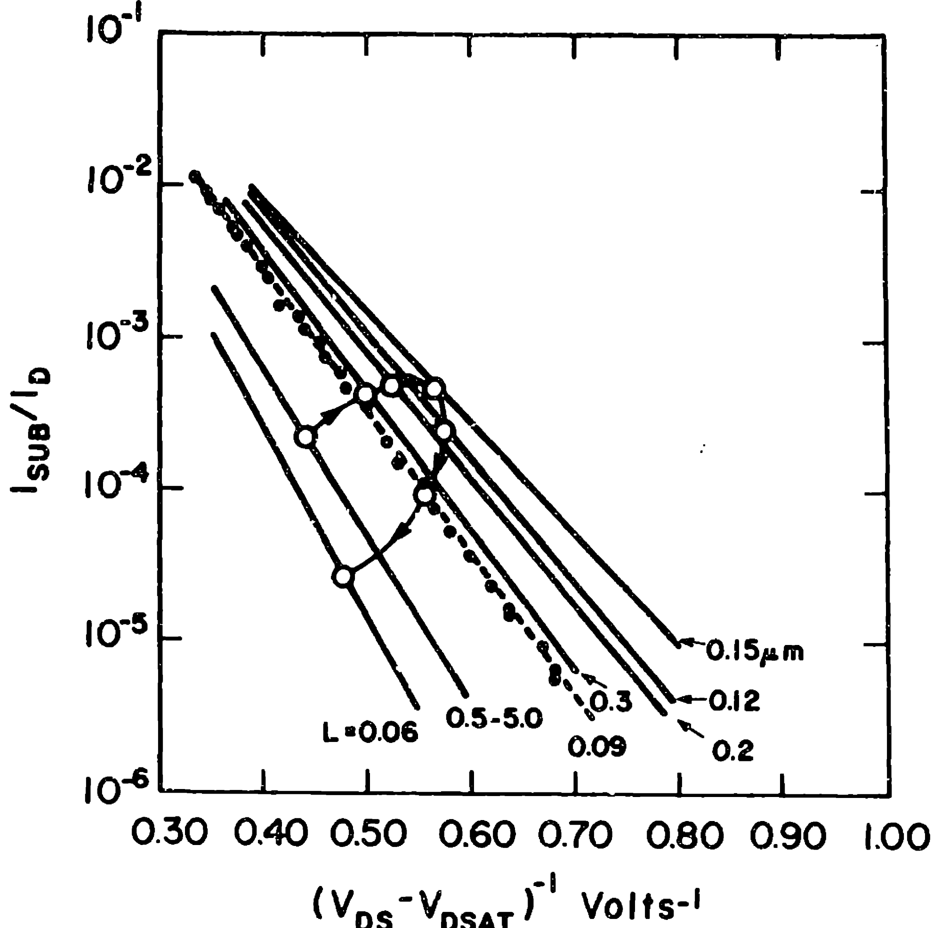


Figure 6.3: Plots of $log(I_{SUB}/I_D)$ vs. $(V_{DS} - V_{DSAT})^{-1}$ at room temperature for devices with channel lengths from 0.06 μm to 5 μm. The arrows trace the "evolution" of the plots as channel length is decreased.

As expected, at a given $(V_{GS} - V_T)$, V_{DSAT} decreases as L is reduced [Sodini]. In the plots of Fig. 6.3 individual data points for a device of $L = 0.09\mu m$ are given to show the typical degree of linearity and data spread.

The behavior of I_{SUB}/I_D vs. $(V_{DS} - V_{DSAT})^{-1}$ for long channel devices $(L > 0.5\mu m)$ at room temperature is as expected. According to the lucky-electron model (on which method of Fig. 6.3 is based) the generated substrate current is exponentialy proportional to the maximum field in the channel or field at the drain edge. Moreover, as it was shown in the previous section, using a simplified two dimensional model of a long channel MOSFET, the maximum field in the channel, can be related to the difference between the drain-to-source voltage and the drain saturation voltage (voltage at which the carrier velocity near the drain saturates), $(V_{DS} - V_{DSAT})$, independent of channel length. Therefore according to the lucky electron model, a plot of $log(I_{SUB}/I_D)$ vs $(V_{DS} - V_{DSAT})^{-1}$ should be a straight line, independent of channel length. This is what is observed for devices with L$> 0.5\mu m$.

As the channel length is decreased below $0.5\mu m$, the plots of $log(I_{SUB}/I_D)$ vs $(V_{DS} - V_{DSAT})^{-1}$ become dependent on channel length, shifting upwards and displaying a decreasing slope. In other words, I_{SUB}/I_D for a given $(V_{DS} - V_{DSAT})$ increases with decreasing L. Indeed, using the two-dimensional device simulator, MINIMOS, it is verified that in our device configuration as L is reduced below $0.5\mu m$ both the maximum field in the channel, E_m, and the width of the (near peak) high field region $(E > 2 \times 10^5\ V/cm)$, at a given $(V_{DS} - V_{DSAT})^{-1}$, increase. The result is shown in Fig. 6.4 (V_{DSAT} was found using the formula (2.5.9) in Chap. 2). It is then reasonable to expect that such an increase in E_m causes the increase in substrate current as shown in Fig. 6.3.

However the measured trend of increasing I_{SUB}/I_D with decreasing L is reversed for

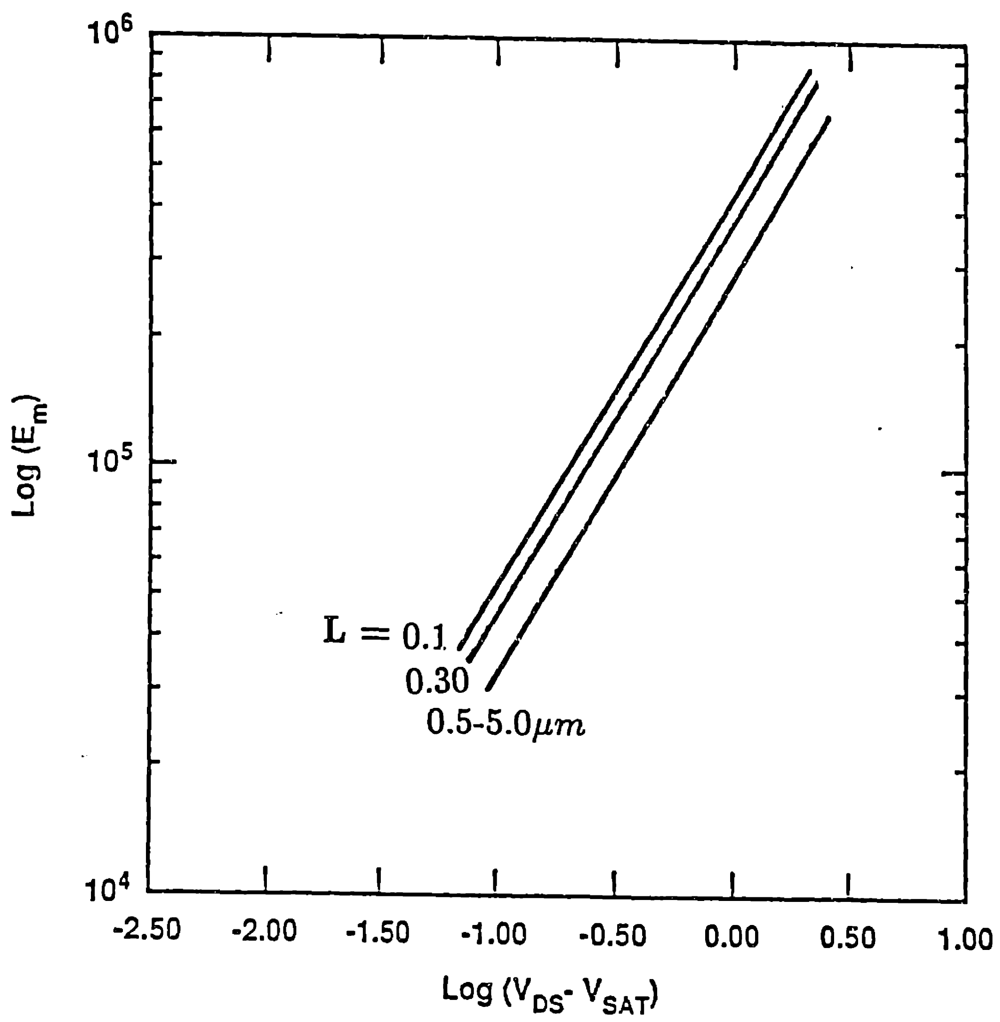


Figure 6.4: Plots of E_m, the maximum field in the cahannel vs. $(V_{DS} - V_{DSAT})^{-1}$ at room temperature for devices with channel lengths from 0.1 μm to 5 μm (As calculated by MINIMOS).

L below about $0.15\mu m$. As can be seen in Fig. 6.3, the $log(I_{SUB}/I_D)$ vs $(V_{DS}-V_{DSAT})^{-1}$ plots start shifting downwards even though E_m is expected to continue to increase.

We believe that it is plausible that because of the finite energy relaxation time and very high fields and field gradients in a very short channel device, the electron temperature, T_e, never reaches its maximum steady state value. Therefore T_e may be actually reduced compared to longer channel devices at the same $(V_{DS} - V_{DSAT})$ value, i.e. there is less interaction between electrons and the lattice. Reduced electron temperature leads to velocity overshoot, and we suspect that it may also be responsible for the reduction in substrate current generation. In fact recent simulation in IBM indicates that velocity overshoot leads to reduction in carriere energy.

Increased carrier velocity due to overshoot in itself can also contribute to the I_{SUB} reduction. In devices where overshoot occurs, the increased carrier velocity at the drain end of channel, where the field is maximum and most of the impact ionization takes place, can cause the carrier density to decrease relative to the case of no velocity overshoot. A reduction of the density of electrons in a high field region can contribute to a further reduction of impact ionization.

Using the scheme developed in Chap. 4, and MINIMOS, the maximum carrier temperature in the channel was calculated, for the cases with and without velocity overshoot. The result is plotted in Fig. 6.5. The lateral field was obtained from MINIMOS. Velocity overshoot leads to lower lateral field in the channel, and a smaller rise is observed as the channel length is reduced. Of course the temperature calculation is not self consistant, and this is why a reduction in the carrier temp[erature can not be observed at small channel lengths. A self consistant calculation leads to lower drain field [Odeh].

Using MINIMOS we have verified that in all devices, even the ones in punchthrough,

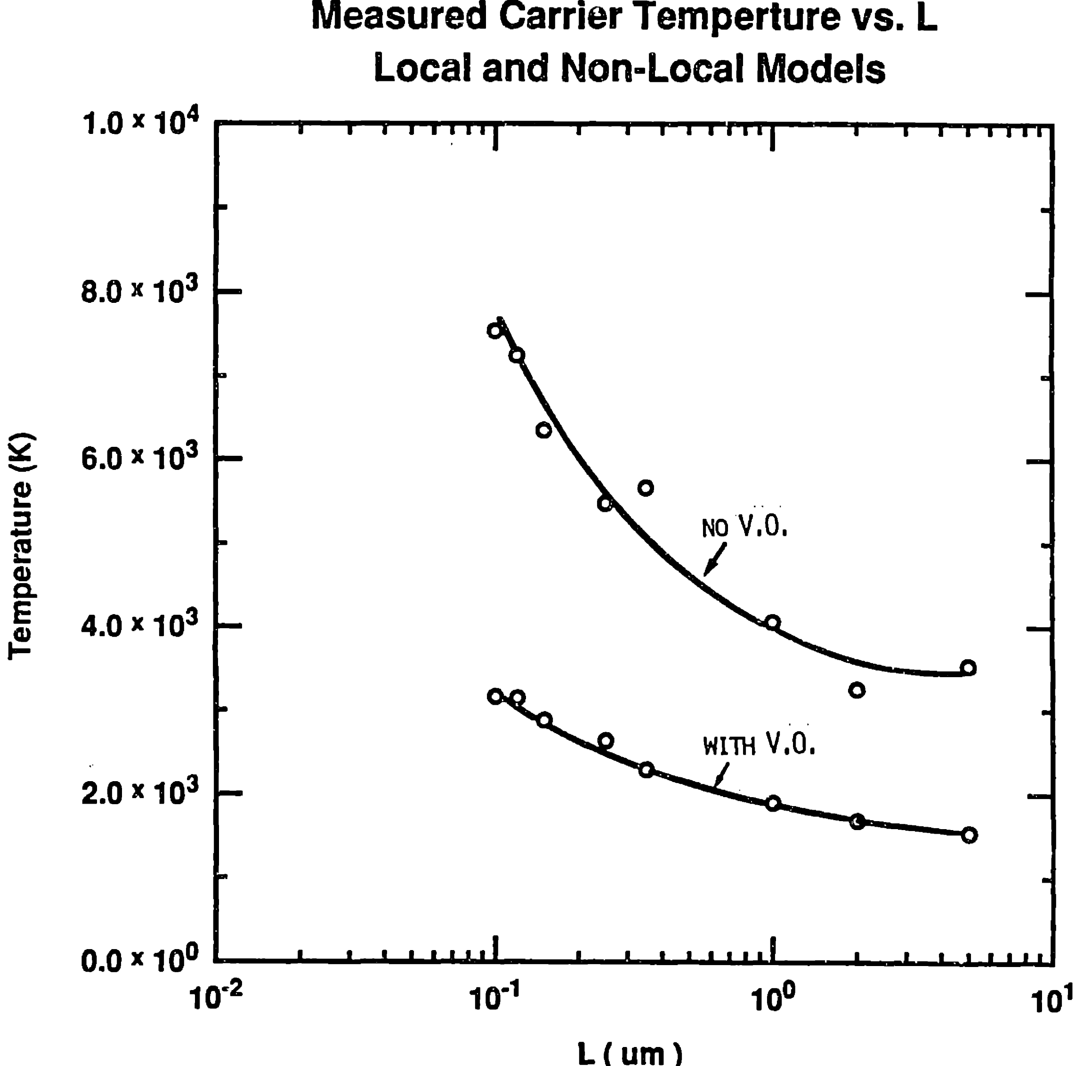


Figure 6.5: The maximum carrier temperauture in the channel at V_{DS} = 2.0 V and V_{GS} = 1.5 V, at room temperature for devices with channel lengths from 0.08 μm to 5 μm with and without velocity overshoot. (Lateral field was calculated by MINIMOS and the temperature was calculated using one dimensional hydrodynamic equations).

current passes through high field region. Furthermore the reduction in I_{SUB}/I_D at a given $(V_{DS} - V_{DSAT})^{-1}$ for $L < 0.15$ μm, was observed in a wide varaity of devices, irrespective of short channel behavior (drain induced barrier lowering or punchthrough). The amount of reduction for $L < 0.15$ μm and increase for $L > 0.5$ μm was critically dependent on channel doping profile. In Chap. 7, reduction insubstrate current in Indium doped devices are presented that are perfectly well behaved and out of punchthrough.

6.6 Experimental Results at 77 K

Substrate currents were also measured in short-channel devices at 77 K. The most remarkable phenomenon is an increased reduction, with respect to 300 K, in the magnitude of substrate current for short-channel devices as compared to long channel devices, under identical bias conditions. Figure 6.5 demonstrates this effect clearly: drain and substrate currents were measured at the same $V_{GS} - V_T$ for a long- and a short-channel MOSFET at room temperature and at 77 K (V_{GS} is the gate bias and V_T is the linear-region-extrapolated threshold voltage). Note the enhanced reduction of the substrate current generation at low temperature for short-channel devices.

For the 2.5 μm channel-length device, at large V_{DS}, both drain and substrate currents increase from 300 K to 77 K. There exists a crossover voltage below which I_{SUB} actually decreases with temperature. This effect has been discussed by [Henning]. However, in the 0.14 μm channel-length MOSFET, I_{SUB} decreases at 77 K for all values of V_{DS}, while I_D increases. It appears that the crossover voltage has moved beyond the measurement range. Figure 6.7 shows this effect more systematically through plots of I_{SUB}/I_D vs. $(V_{DS} - V_{DSAT})^{-1}$ at 77 K and at room temperature for devices of different L. For long channel devices, there is a drop of I_{SUB}/I_D at high $(V_{DS} - V_{DSAT})^{-1}$, i.e. low V_{DS}, and an increase of I_{SUB}/I_D at low $(V_{DS} - V_{DSAT})^{-1}$, as also previously reported by [Lau]. For

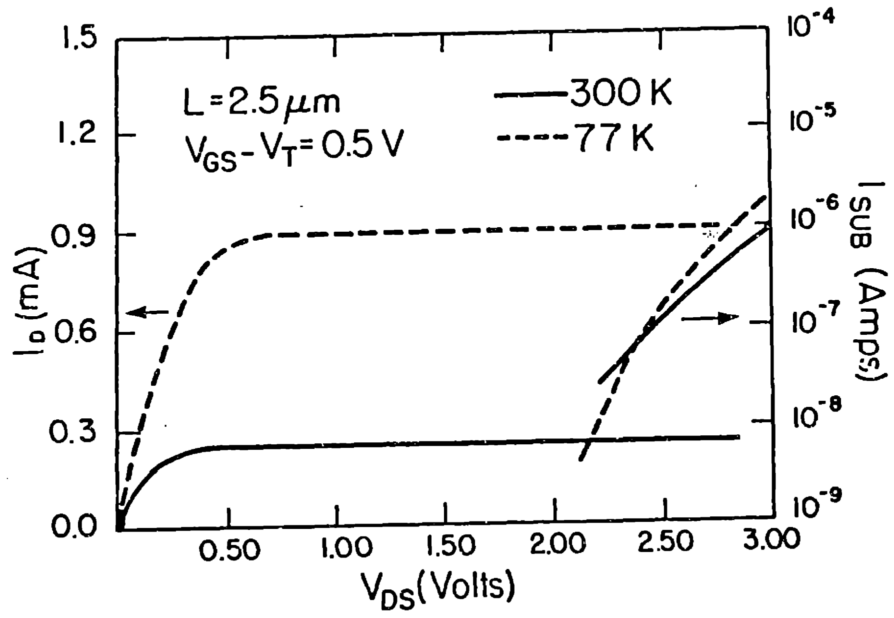


Figure 6.6a: I_D and I_{SUB} vs. V_{DS} at $(V_{GS} - V_T) = 0.5V$ for a long-channel device, L= 2.5μm (V_T = 0.19 V, at 300 K and V_T = 0.46 V, at 77 K), at room temperature (solid line) and at 77 K (dashed line).

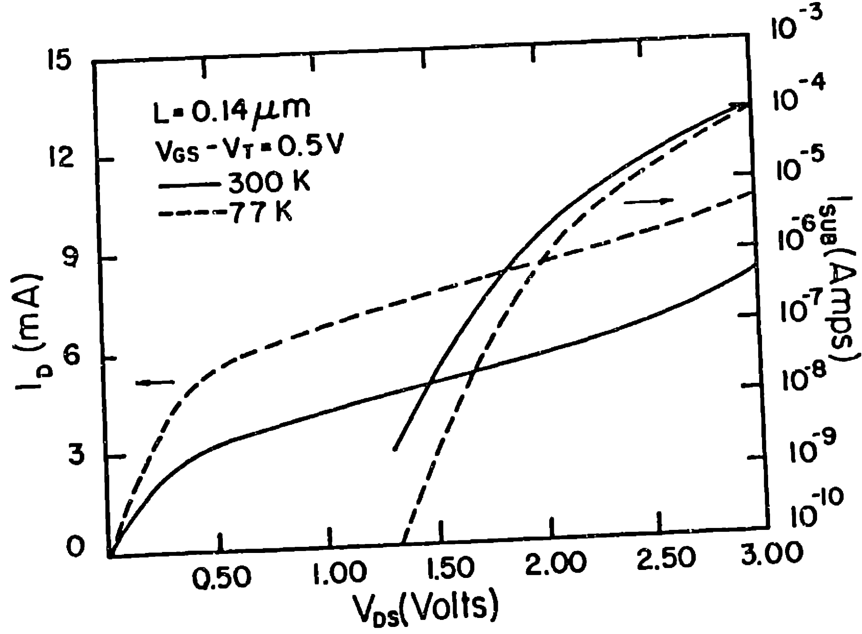


Figure 6.6b: I_D and I_{SUB} vs. V_{DS} at $(V_{GS} - V_T) = 0.5V$ for a short-channel device, L= 0.14μm (V_T = 0.07 V, at 300 K and V_T = 0.33 V, at 77 K).

short-channel devices, however, I_{SUB} is reduced more from room temperature to 77 K and in the range of our measurement, $0.35\ V^{-1} < (V_{DS} - V_{DSAT})^{-1} < 0.8\ V^{-1}$ no increase in I_{SUB}/I_D was observed. Because of the change in slope, it may be surmised that the crossover between the plots for 300 K and 77 K has shifted to larger drain biases, beyond the measurement range, as noted earlier.

The results at 77 K are more surprising than the result at 300 K. The reduction of substrate current at low V_{DS} suggests a reduction in effective impact ionization coefficient which is contrary to the calculation of [Baraff] and the measurements of [Crowell] which give an increase in impact ionization coefficient at 77 K. At 77 K, the optical phonon emission length is longer and this is expected to lead to higher impact ionization coeffecient (6.2.1). Perhaps non-local effects resulting from finite energy relaxation time in the region of abruptly changing electric field are responsible for both long channel and short channel device behavior, and these effects become more noticeable at low temperature and shorter channel lengths.

The unexpected reduction in the substrate current, which is tentatively attributed to non-stationary effects, highlights the need to include the higher moments of BE in calculations of lateral field and impact ionization of MOSFETs.

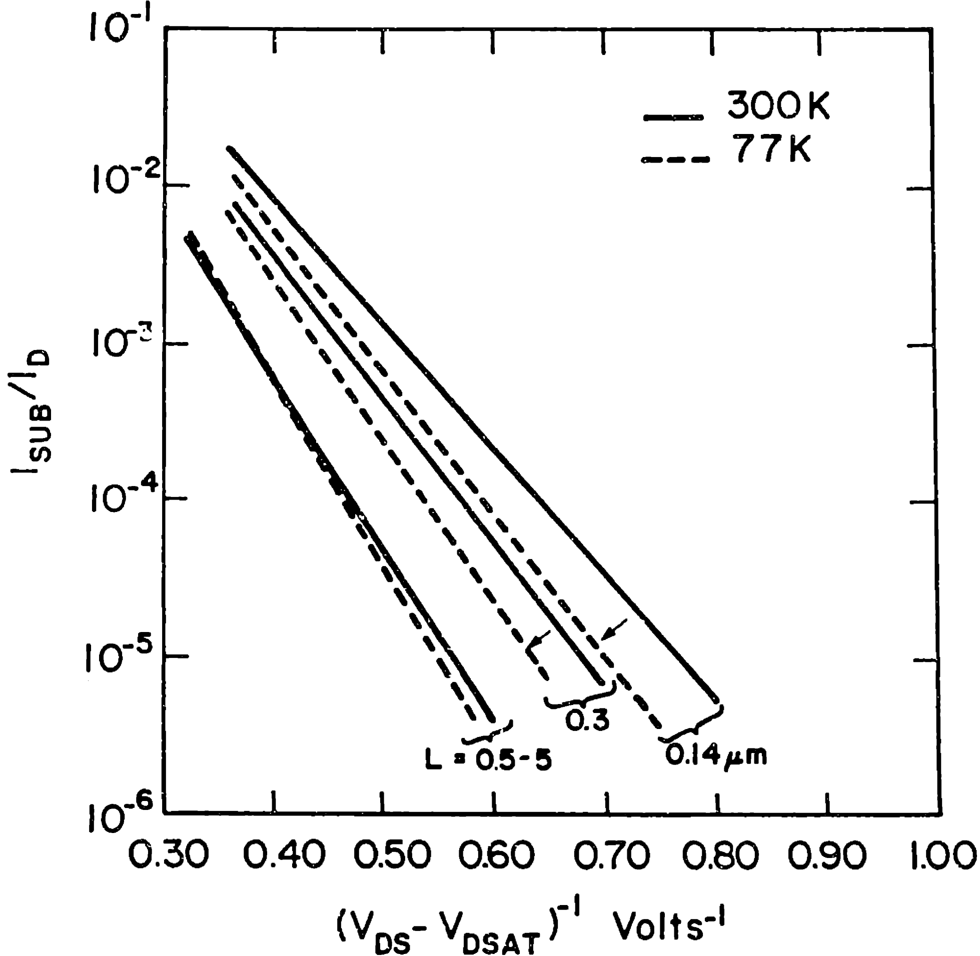


Figure 6.7: Plots of $log(I_{SUB}/I_D)$ vs. $(V_{DS} - V_{DSAT})^{-1}$ for $0.14\mu m < L < 5\mu m$, at room temperature and 77 K. Note the enhanced reduction of the substrate current generation at low temperature for short-channel devices.

Chapter 7

Indium Implanted MOSFET Channel

7.1: Introduction

As device dimensions are reduced into the deep submicron range, substrate doping must be increased to reduce unwanted short-channel behavior. Uniform heavy substrate doping results in reduced channel mobility, increased hot electron effects, and increased body effect. To improve channel mobility (and thus gain), non-uniform channel doping has been investigated [Shahidi]: the surface doping is kept low to reduce impurity scattering, but below the surface, there is heavy p doping to reduce short channel effects. This was initially achieved by boron implantion (at 50 keV) and minimization of high temperature steps after the implant. In such devices electron velocity overshoot at room and liquid nitrogen temperatures was observed.

In an attempt to make the surface as lightly doped as possible, and bring the peak of substrate doping as close to the surface as possible, indium was used as an alternative

channel implant. Indium has an atomic weight of 115, and the implant peak can be placed within 10's of nm of the surface at moderate to high implant energies. The profile is skewed toward the bulk, resulting in very low surface concentration. Deep-submicron-channel-length devices were fabricated using x-ray lithography and indium channel doping, and were compared to boron implanted devices.

Figure 7.1 compares SUPREM simulated dopant concentration at three different indium implant energies with one boron implant, all at a dose of 5×10^{12} cm^{-2}, after our normal high temperature steps. Since indium has a large segragation coefficient into the SiO_2, the surface has very low dopant concentration [Antoniadis]. Previously, this indium's high segragation into the SiO_2 was mentioned as an obstacle to its use as a viable alternative to boron in the field region in MOS technology [Antoniadis]. In this work, it is used as the channel implant, and highly non-uniform doping is desired. Because of its high atomic number, it is possible to place the peak very close to the surface and achieve this highly non-uniform doping.

Initial MINIMOS simulations, indicate the superiority of indium implanted devices over the boron implanted (uniform heavy doping) ones in terms of drain induced barrier lowering. Figure 7.2 shows the threshold reduction as a function of the V_{DS}, for a 0.08 μm device, for channel implanted with indium and uniform high achieved with boron (All the implant doses are the same). It can be observed that for low energy indium implant, the threshold fall is less than that of the uniformly high implanted channel. Furthermore to achieve the same threshold, the indium dose must be increased. This

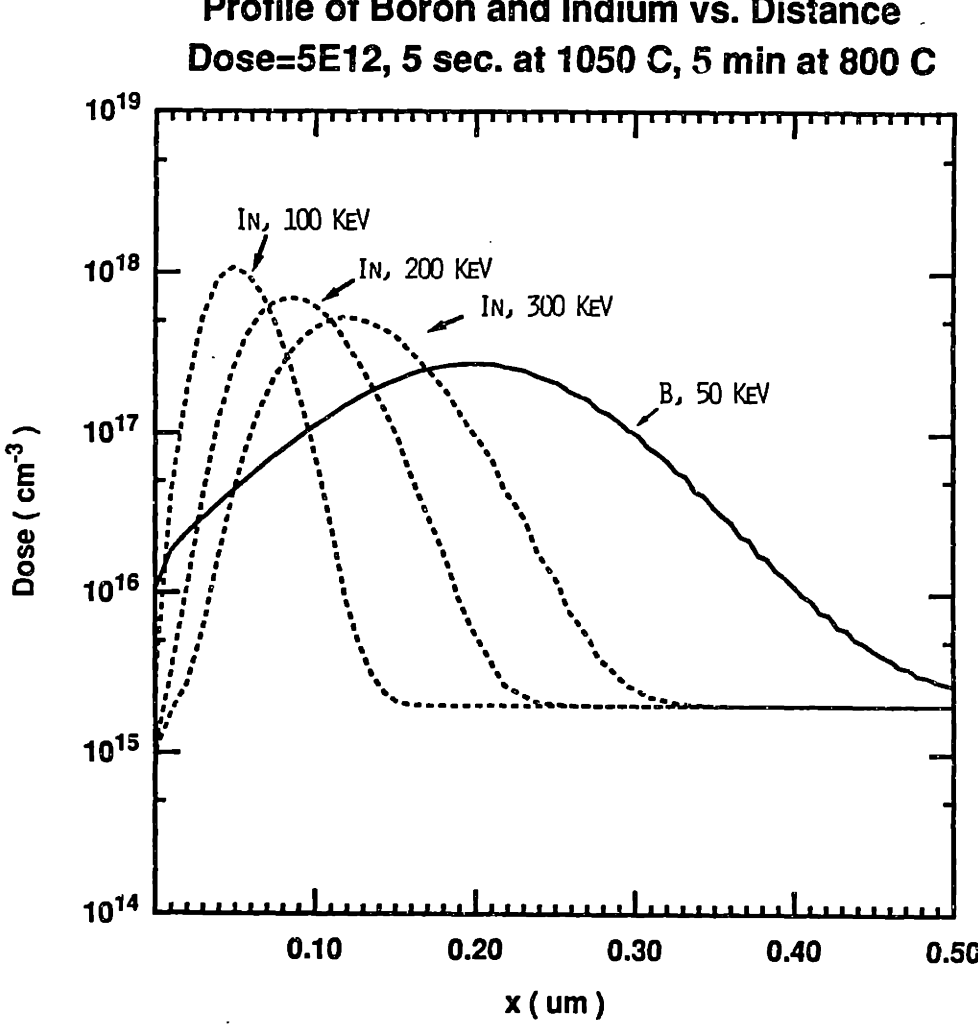


Figure 7.1: Concentration of Indium implanted at three different energies and boron at 50 KeV, as simulated by SUPREM.

would result in further reduction in V_T roll off.

7.2 Experimental Results

Short channel devices were fabricated using indium implanted channels. Indium energies of 100, 200 and 300 Kev were used. The result were compared to a born implanted channel of the same dose and at used before).

7.2.1 Necessity of Deep Implant

Indium alone produces a heavy doping just below the surface. This results in high threshold and reduced short channel effect. But at low indium implant energies, and short channel lengths, bulk punchthrough occurs, i.e. a conduction path between the source and drain through the bulk is formed. Figure 7.3a shows the drain current as a function of V_{GS} for two different V_{DS}. For high V_{DS}, device turn off is not possible. To solve this problem with the device, a second deep implant is placed under ther channel (Boron, 1E13, 180 KeV, so that surface doping is not affected by this second implant). The second implant effectively blocks the bulk conduction. This is demonstrated in Figure 7.3b (The small curent for low V_{GS} is due to gate current). In most of this work, the second deep implant was always done to stop the bulk punchthrough.

7.2.2 Mobility

Mobility was measured in indium doped samples. Because of the lower surface doping, high mobility was expected. Electron mobility in Si MOSFETs is a strong

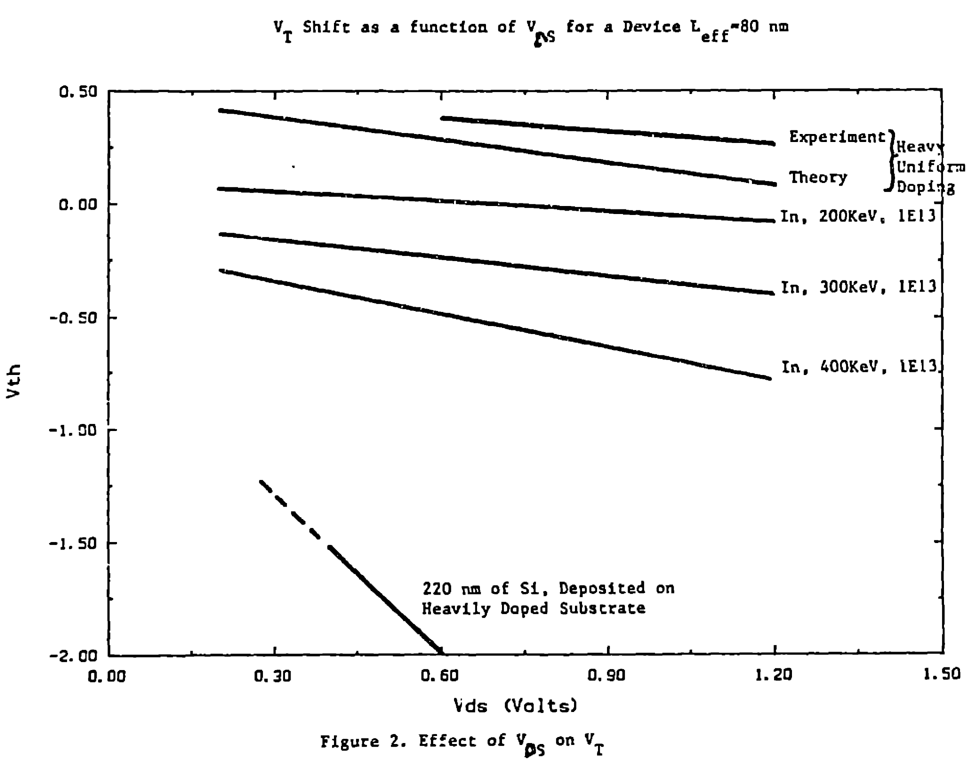


Figure 7.2: Reduction in V_T as function of V_{DS} for different values of implant energy.

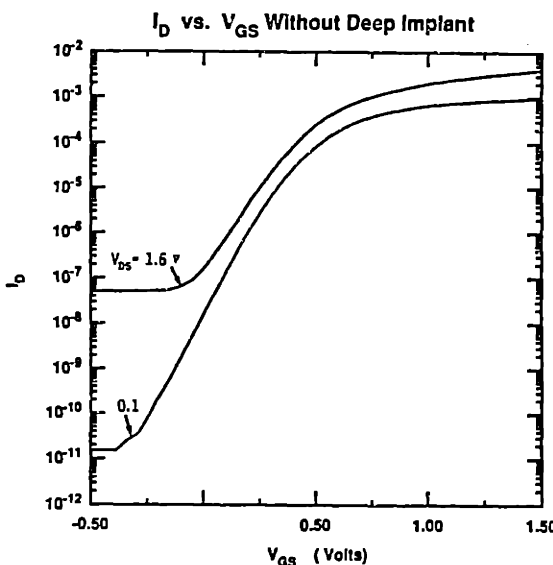


Figure 7.3a: Subthreshold conduction for a submicron MOSFET without the deep implant.

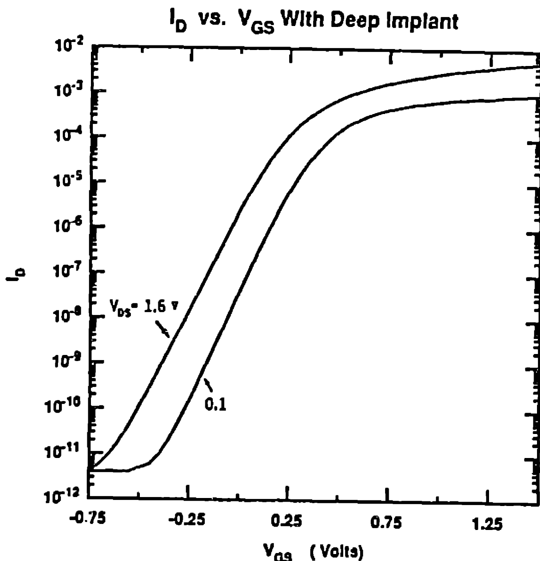


Figure 7.3b: Subthreshold conduction for a submicron MOSFET with the deep implant.

function of the normal field at the threshold. It has to be emphasised that no systematic mobility measurement was done to seperate the effect of the normal field and the high substrate doping on mobility. Figure 7.4 summarizes our results. Mobility was measured for two extremes: One is a boron doped channel (50 KeV, 5E12, V_T=0.15 V), and the other sample is indium doped with twice the dose and the minimum energy and much higher threshold voltage (100 KeV, 1E13, V_T=0.6 V). As can be seen, the increase in the threshold and improvement in short channel behavior, caused only 10% reduction in mobility (from 420 to 380 $cm^2/V.s$). It has to be emphasised that in these samples the gate was deposited by e-beam lithography, and some damage to gate oxide migth have occured (Lower mobility was obtained).

7.2.3 Device Characteristics

As expected, at low indium implant energies one achieves high device threshold in nMOSFET's. Turn off charactristics were excellent. Table 7.1 summarises the device characteristics and compares them to the boron doped channel:

In summary indium implantation improves in short channel behavior, (At a small cost to mobility). In indium doped samples, an increase in body effect is observed as expected. Figure 7.5a compares the body effect of two cases of boron and indium doped channel as in table 7.1. As expected, boron has a much smaller body effect. Reducing the implant dose, lowers the body effect (Figure 7.5b). Reducing the indium implant energy has little effect on body effect (Figure 7.5c).

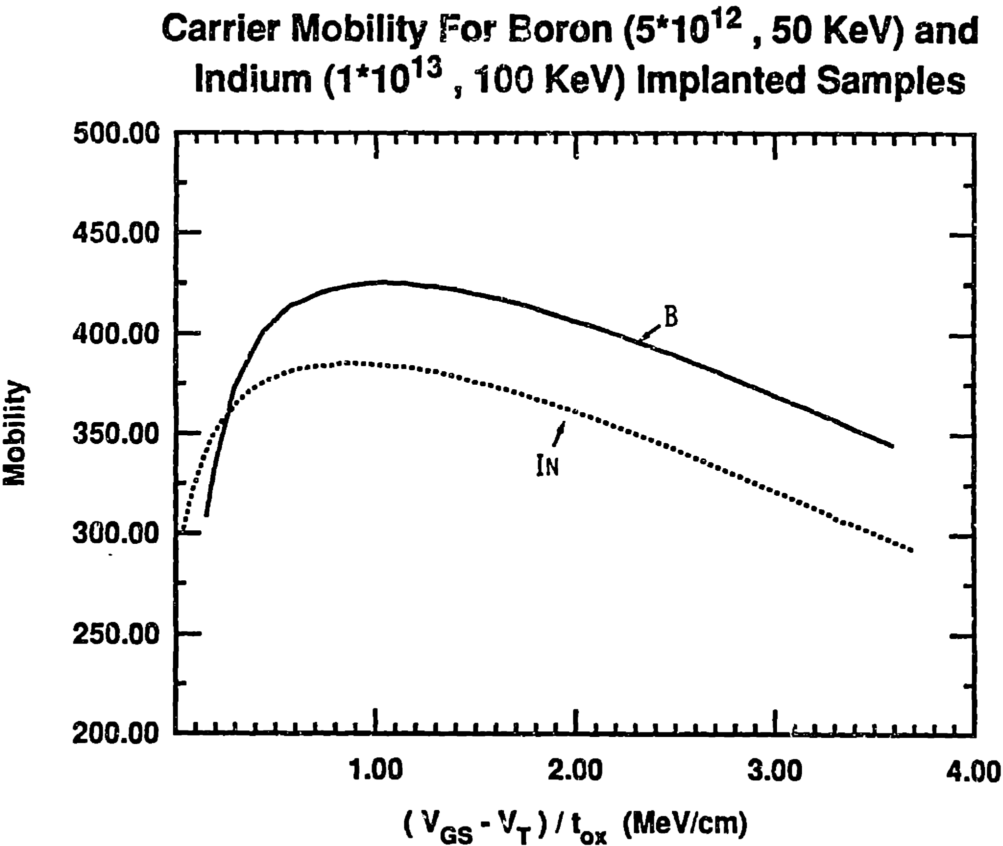


Figure 7.4: Mobility in a boron doped channel and an Indium doped channel (with twice the dose and much shallower).

	Boron 50 KeV, 5E12			Indium 100 KeV, 1E13		
L (um)	5	0.15	0.1	5	0.15	0.1
V_T (volt)	0.12	-0.13	-0.26	0.6	0.33	0.24
Mobility	420	-	-	380	-	-
S * (V_{DS}=0.1 v)	100	180	400	85	85	110
S * (V_{DS}=1.6 v)	100	350	1000	85	85	175
V_{EFF} (*10^7 cm/s)		0.95	1.1		0.85	0.95

* S is subthreshold slope (mV/dec)

t_{ox} is 5.3 nm.

Table 7.1: Comparision of indium and boron doped channels

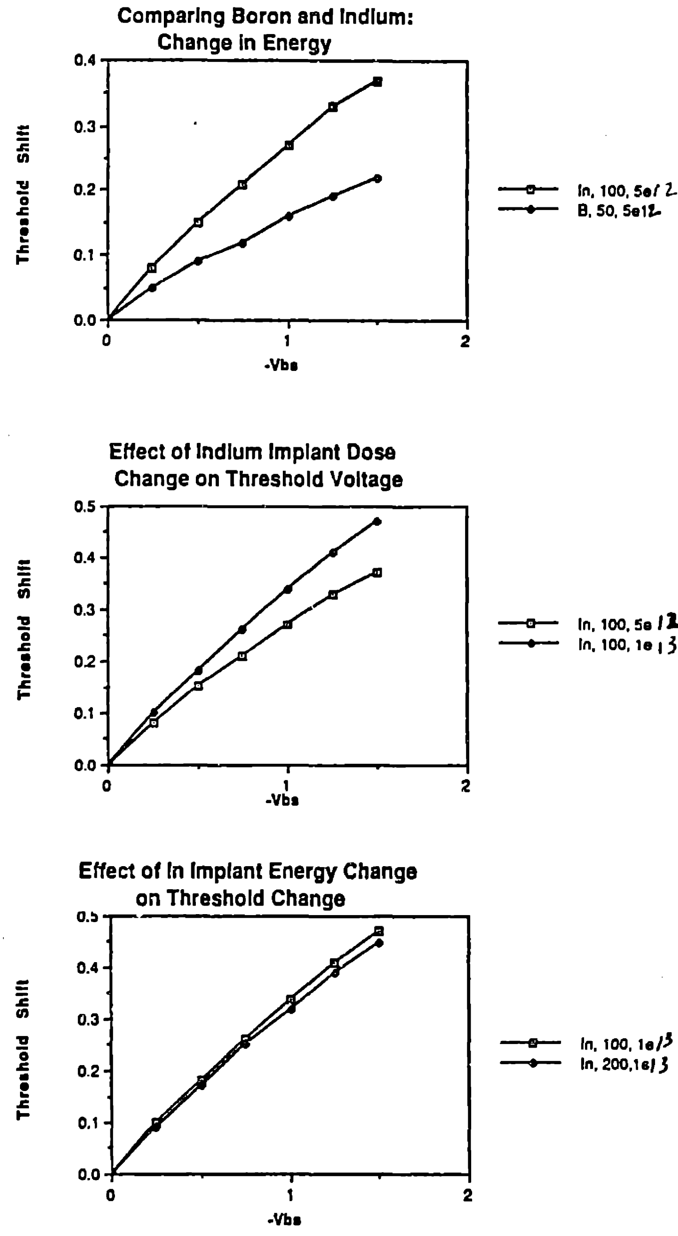


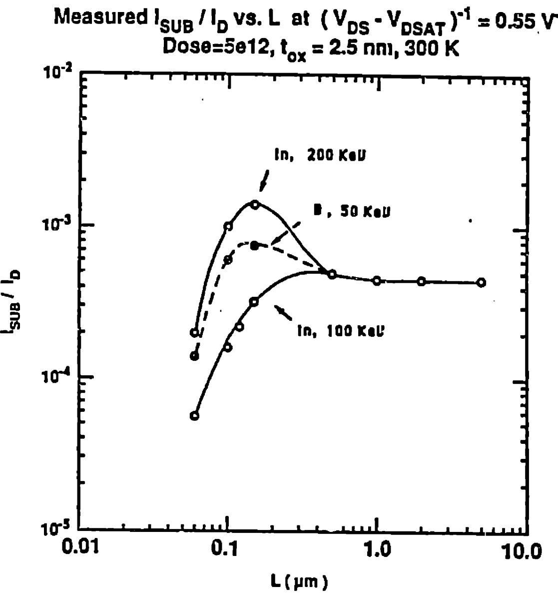
Figure 7.5: (a) Comparision of Inium and boron doped samples (Table 7.1) in terms of body effect, (b) Effect of dose on body effect, and (c) Effect of implanr energy on body effect ($t_{ox} = 5.3nm$)

7.2.4 Hot Electron Behavior

Hot electron generated substrate currents were measured in samples without and with the deep born implant. In all these samples, substrate current reduction was observed for $L < 0.15$ μm. The amount of reduction was a strong function of the indium implant energy and the deep boron doping. Previously, an increase in normalised substrate current ,I_{SUB}/I_D , for devices with $L < 0.5$ μm because of the increase in the maximum field in the channel, and a reduction in I_{SUB}/I_D for $L < 0.15$ μm due to non-equilibrium of electron temperature were reported. In Fig. 7.6a, I_{SUB}/I_D vs. channel length at a constant $(V_{DS} - V_{DSAT})^{-1}$ is plotted. The strong dependece of substrate current generation on indium implant energy for short-channel devices is evident. Furtheremore, for the shallowest implant (100 KeV) no increase in I_{SUB}/I_D for $L < 0.5\mu m$ was observed, only the reduction in normalized I_{SUB} for devices with L $< 0.15\mu m$ is observed.

Figure 7.6b shows the same measurement in samples with the heavy deep p^+ doping. In this case substrate current generation was independent of the indium implant energy. This is probably indicayes that the heavy deep p^+ doping dominates the field behavior at the drain edge.

To prove that the reduction in the substrate current is independent of the punch-through behavior of the device, drain and the corresponding substrate currents of two short channel devices (L= 0.1 and 0.15 μm) are presented in Fig. 7.7. These devices have subthreshold slopes of 85 and 175 mV/dec at $v_{DS} = 1.6$ V. Figure 7.7a shows



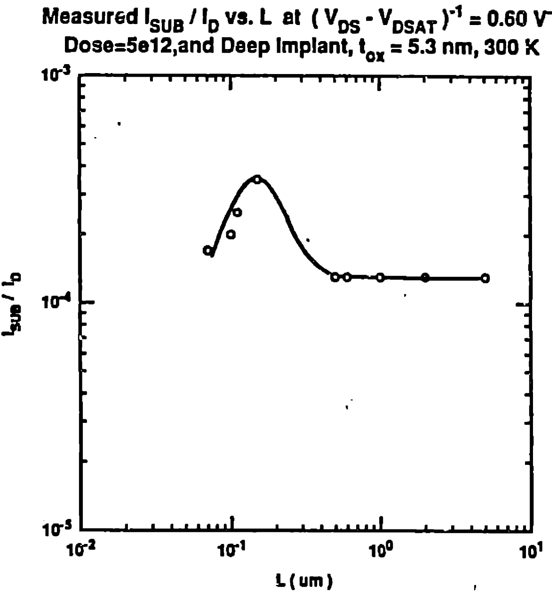
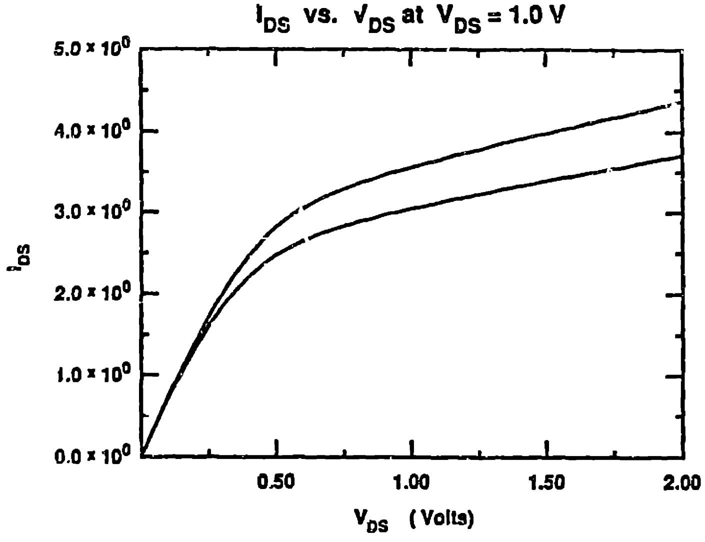


Figure 7.6: Plots of $log(I_{SUB}/I_D)$ vs. channel length at $(V_{DS} - V_{DSAT})^{-1} = 0.55$ V^{-1} for different values of implant energies. Note the strong dependence of substrate current generation on implant energy for submicron devices. (a) is without the deep boron implant, and (b) is with the deep implant.

the drain current at $V_{GS} = 1.2$ V, and Fig. 7.7b shows the corresponding substrate currents. Note the reduction in the substrate current, despite the increase in the drain current.

In conclusion, it is shown that indium can be used as an alternative channel implant species. Indium implanted devices have superior short channel behavior.



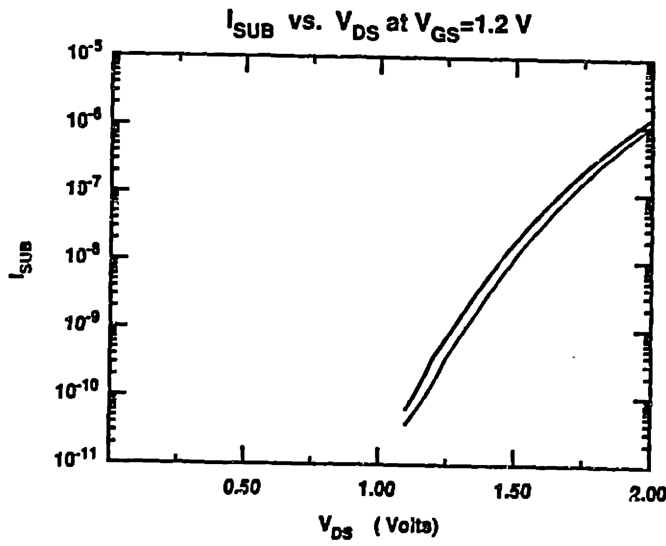


Figure 7.7: The drain and the corresponding substrate current for two short channel devices (L=0.1 and 0.15 μm), at $V_{GS} = 1.2$ V. Note the reduction in the substrate current, despite the increase in I_D.

Chapter 8

Conclusion

This thesis is a study of non-stationary effects in deep sub-micron channel MOSFETs. These are effects that are caused by the lack of equilibrium between the carriers and a high and abruptly varying electric field. To be able to observe these effects clearly, the scattering times along the drift direction of carriers have to be maximized.

The basic tool used in this study was deep submicron channel MOSFET: It has a high field region in the channel and a source of cold carriers in the source. Carriers are injected from the source into the high field channel. If the device is sufficiently short, and the lateral electric field in the channel is sufficiently high, non-stationary effects can be observed. In order to observe these effects clearly at higher temperatures, carrier scattering in the channel must be reduced; i.e. the channel mobility must be high. The first part of this thesis dealt with the fabrication of high mobility deep-submicron channel MOSFETs. The basic process used in fabrication of short channel MOSFETs already existed, and some slight modifications were used to obtain devices with higher mobility and smaller parasitic resistances. In our MOSFETs, high mobility was achieved by a highly non-uniform channel doping: The surface doping was kept

low to increase the mobility. A high dopant region was placed just below the surface to prevent punchthrough. This non-uniform doping was initially achieved by means of boron implantation. The non-uniform doping allowed us to achieve mobilities of up to 450 $cm^2/V.s.$

Once the high mobility MOSFETs were fabricated, velocity overshoot were readily observed at room temperature and 77 K for the first time. That is, effective carrier velocities in excess of 10^7 cm/s were measured in devices with L$<$ 0.1 μm at room temperature. This observation proves the existence of non-stationary transport. The magnitude of such an effect is small at room temperature (20% over the bulk value) and high at 77 K (80% over the bulk value). It is not clear if a complicated alteration of simple drift-diffusion is necessary for modeling of devices operating in this regime.

Velocity overshoot is caused by reduced carrier temperature in the channel (relative to the one corresponding to the local field). This prompted investigation of hot electron effects in the deep submicron range. Specifically,the hot-electron-generated substrate current was measured, and a reduction in the substrate current was observed for devices with channel length L$<$ 0.15 μm. An increased amount of substrate current reduction was observed at 77 K in short channel devices. It is proposed that a reduction of carrier temperature due to abrupt variation in the lateral electric field and a reduction in the carrier density in the high field region because of velocity overshoot, are responsible for the reduction in the substrate current generation in the deep submicron channel length range. The substrate current is reduced by an order of mag-

nitude in very short channel devices. This highlights the need for incorporation of a more complete carrier transport model when studying the hot electron effects in deep submicron range at room temperature and/or 77 K.

To reduce short channel effects in deep submicron channel devices, indium was used as a channel implant for the first time. It is possible to keep the surface doping low and bring the heavy p^+ doping very close to the surface, thus reducing the short channel effects. At the same time, the reduction in substrate current was observed was observed in well behaved devices. Velocity overshoot was observed in indium implanted devices too.

These observations, prove the exictance of the so called non-stationary effects, and highlight the need for their inclusion (through higher moments of Boltzmann Equation or use of Monte Carlo method) in modeling of short channel devices.

Appendix A

Lateral Channel Field [Al-Mansy]

The first attempt to model the lateral electric field in the drain region, in the saturation region of operation, is due to [El-Mansy]. He assumes that the drift is the dominant form of the transport:

$$I_D = -W\mu_e Q_{MOB}\frac{dV_S}{dy} \tag{A.1.1}$$

where Q_{MOB} is the mobile charge density, V_S is the surface potential, and μ_e is the effective electron mobility. For electron mobility, in a region of high lateral field, El-Mansy uses:

$$\mu_e = \frac{\mu_o}{1 + \frac{\mu_o}{v_{SAT}}\frac{dV_S}{dy}} \tag{A.1.2}$$

Integrating the expression from source to drain, one obtaines:

$$I_D = \frac{\mu_o W}{L_e}\int_{V_{SB}}^{V_{DB}} Q_{MOB} dV_S \tag{A.1.3}$$

where L_e is the effective channel length as defined by:

$$L_e = L(1 + \frac{\mu_o}{v_{SAT}}\frac{V_{DS}}{L}) \tag{A.1.4}$$

i.e. effect of velocity saturation on the current is the same as increasing the channel length ,L, by amount proportional to the V_{DS}.

To find the electric field, El-Mansy models the drain region of the device, i.e.region in which the gradual channel approximation breaks down:

$$\frac{\partial V(x,y)}{\partial x}|_S = K.\frac{\partial V_S}{\partial y} \qquad (A.1.5)$$

where $V(x,y)$ is the channel potential, S indicates the value at the surface and K is some large number. For the value of the surface potential at which the GCA breaks down, V_{S0}, El-Mansy uses:

$$V_{S0} = \frac{V_{DSAT0}}{1 + F.(t_{ox}/L)} \qquad (A.1.6)$$

where V_{DSAT0} is the surface potential for pinchoff (zero mobile charge), and F is a slow varying function of gate bias. Notice that this expression for the value of the surface potential seperating the source and drain regions is similair to V_{DSAT} derived in Chap. 2.

The drain region is modeled as a rectangle in which the boundaries are given as follows (Figure A.1): Boundary 1 seperates source and drain regions: $V_S = V_{S0}$. Boundary 2 and 3 are the oxide-silicon and the junction interface. Boundary 4 is at a distance W_{depl} from the surface. At this boundary, both the potential and the electric field are assumed to be zero.

Applying the Gauss' Theorem to the boundary of the rectangle shown in Fig. A.1:

$$\int D.dS = Q_{Total} \qquad (A.1.7)$$

If E_1 and E_x are the electric fields prependecular to the boundaries 1 and 2, and E_y is the

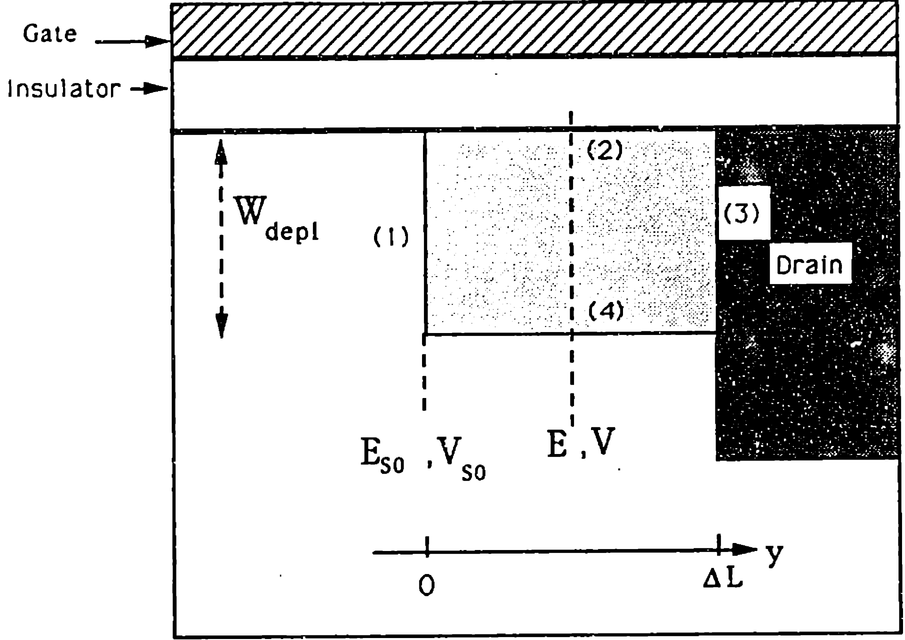


Figure A.1: Drain region of A MOSFET as used by El-Mansy

lateral electric fied at distance y form the saturation point, then:

$$\int D.dS = \int_0^{W_{depl}} (\epsilon_{Si}E_y - \epsilon_{Si}E_1)dx - \int_0^y \epsilon_{Si}E_x dy$$

$$= -q\int_0^y \int_0^{W_{depl}} (N_A + n)dxdy \qquad (A.1.8)$$

where N_A and n are the ionized substrate acceptor density and the mobile charge densities. Using the fact that $E_y = \frac{\partial V_S}{\partial y}$, $E_x = \frac{C_{ox}}{\epsilon_{Si}}(V_{GS} - V_T - V_S)$, and $E_1 = E_{SAT}$ then:

$$\int_0^{W_{depl}} (\frac{\partial V_S}{\partial y} - E_1)dx - \int_0^y \frac{C_{OX}}{\epsilon_{Si}}(V_{GS} - V_T - V_S)dy$$

$$= -\frac{q}{\epsilon_{Si}} \int_0^y (qN_A W_{depl} - Q_{mob})dy \qquad (A.1.9)$$

Now El-Mansy assumes that the average (over x) of the longitudial field is proportional to the value of the longitudial electric field at the surface:

$$\int_0^{W_{depl}} (\frac{\partial V_S}{\partial y})dx = \frac{W_{depl}}{3}\frac{\partial V_S}{\partial y} \qquad (A.1.10)$$

Using these relations and differentiating both sides of the equation (A.1.11) with respect to y, one obtains:

$$\frac{d^2 V_S}{dy^2} = \frac{3C_{ox}}{\epsilon_{Si}W_{depl}}(V_S - V_{GS} - V_T) + \frac{3}{\epsilon_{Si}}(qN_A - \frac{Q_{MOB}}{W_{depl}}) \qquad (A.1.11)$$

Rewritting $\frac{d^2V_S}{dy^2} = \frac{1}{2}\frac{d}{dV_S}(\frac{dV_S}{dy})^2$, and integrating from boundary 1 to drain, one obtains:

$$E_D^2 - E_{SAT}^2 = \frac{3C_{ox}}{\epsilon_{Si}W_{depl}}[(V_{SD} - V_G - V_T)^2 - (V_{S0} - V_G - V_T)^2] + \frac{3qN_A}{\epsilon_{Si}}(V_{SD} - V_{DSAT})$$

$$-\frac{3}{\epsilon_{Si}W_{depl}}\int_{V_{DSAT}}^{V_{DS}} Q_{MOB}dV_S \tag{A.1.12}$$

From the expression for the current, Eq. (A.1.3):

$$\int_{V_{S0}}^{V_{DS}} Q_{MOB}dV_S = -\frac{1}{W\mu_o}(L_eI_D - L_{eSAT}I_{SAT}) \tag{A.1.13}$$

where L_{eSAT} is the effective channel length when V_{S0} is applied to the drain side. Finally for the lateral electric field at the drain, one obtaines:

$$E_D^2 = E_{SAT}^2 + \frac{3C_{ox}}{\epsilon_{Si}W_{depl}}[(V_{SD} - V_G - V_T)^2 + (V_{S0} - V_G - V_T)^2] + \frac{3qN_A}{\epsilon_{Si}}(V_{SD} - V_{S0})$$

$$+\frac{6}{\epsilon_{Si}W_{depl}W\mu_o}(L_eI_D - L_{eSAT}I_{SAT}) \tag{A.1.14}$$

This expression gives the lateral field at the drain side, and at first migth seem to be useless, since it gives the field only at one point. But as will be shown later, the value of the field at the drain side, alone determines the amount of hot electron generation. Equation (A.1.12) for E_D illustrates the contribution of various terms. The first term is the contribution of the field at the saturation point, and is independent of the V_{DS}. The second term, which is usually the dominant one, is the cosequence of the divergence of the carriers into the bulk. This term depends on the V_{DS}, C_{ox}, and substrate doping (through W_{depl}), but is independent of channel length. The third term depends on the substrate doping. The last term has the effect of the velocity saturation. It can be written as:

$$L_eI_D - L_{eSAT}I_{SAT} = L(I_D - I_{DSAT})$$

$$+\frac{\mu_o}{v_{SAT}}[I_D(V_{DS} - V_{S0}) + V_{S0}(I_D - I_{SAT})] \tag{A.1.15}$$

This term shows the contribution of mobile charges and the drift velocity saturation separately. The drift velocity saturation term is inversely proportional to L (through the I_D term) and becomes more significant as the device channel length is reduced. This term is linear in the V_{DS}. Velocity saturation produces a build up of the mobile carriers in the drain section, and this is equivalent to having a drain section with heavier doping.

It can thus be seen that the the major contribution to the drain field is caused by the field, normal to te surface, and the velocity saturation and the mobile charges have little effect on the drain field.

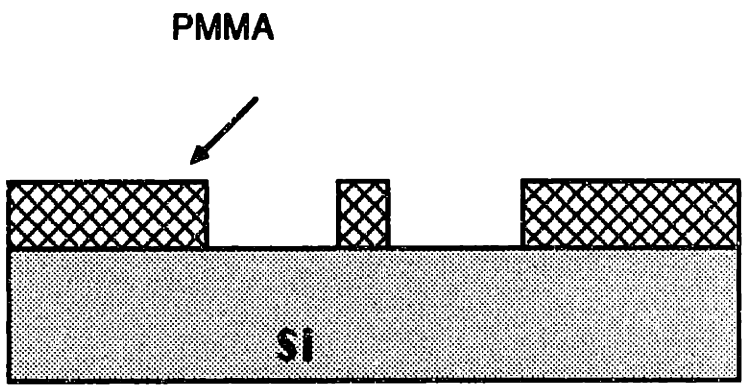
Appendix B

Detailed Processing Steps for Device Fabrication

1. Start with lightly doped p-type <100> wafer.

2. - RCA clean (organic).
   - Spin 0.23 μm of PMMA (Thickness not critical).
   - Bake at180 C for 1 hour.

3. - X-ray exposure.
   - Develop.

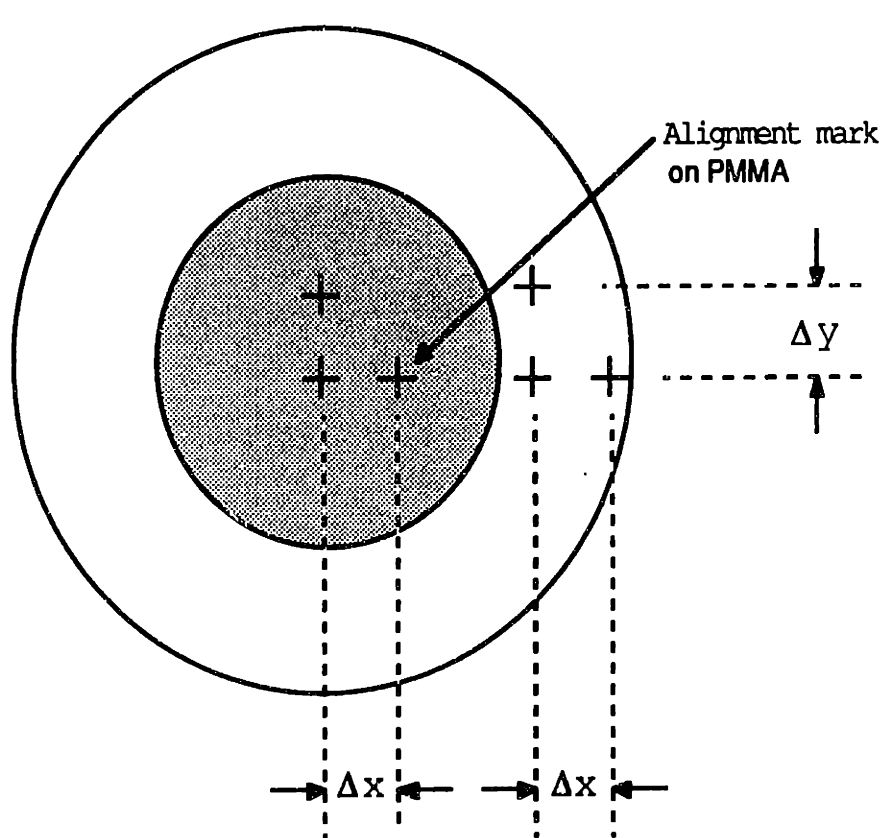


4. Using PMMA as mask, implant source drain, with As: 7E15 at 30 KeV .

5. Spin 1 μm positive resist, soft bake.

6. Using the alignment marks from the PMMA pattern and the GCA stepper, find new stepping distance, and expose new alignment marks on the wafer.

7. Cover the x-ray patterned area with resist, and etch the alignment marks into wafer, using SF6 plasma.
   The pattern on alignment marks on PMMA and wafer covered with resist is as shown below. (Alignment marks are everywhere, but only 6 are shown).



8. Strip all the resist and PMMA in oxygen plasma.

9. Spin 1 μm positive resist, soft bake.
   - Pattern for the deep Phosphorous implant.
   - Hard bake.

10. P implant: 3E15 at 180 KeV.

12. Remove resist in oxygen plasma,

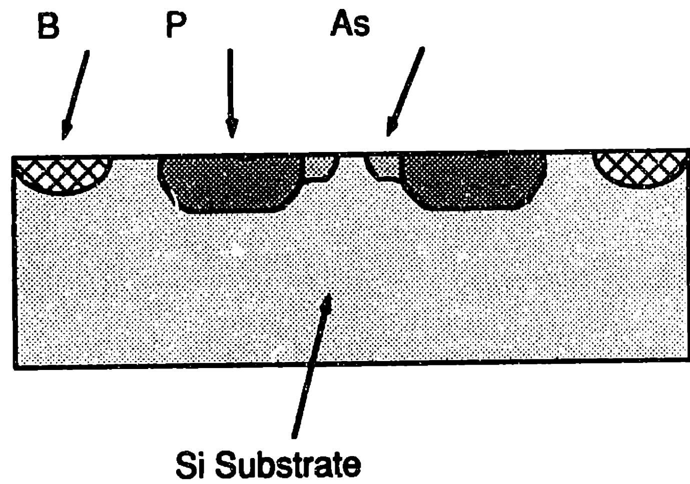
13. -Spin 1 μm positive resist, soft bake.
    -Pattern for the p+ ring around each device.

14. Implant with Boron 1E13 at 20 KeV.

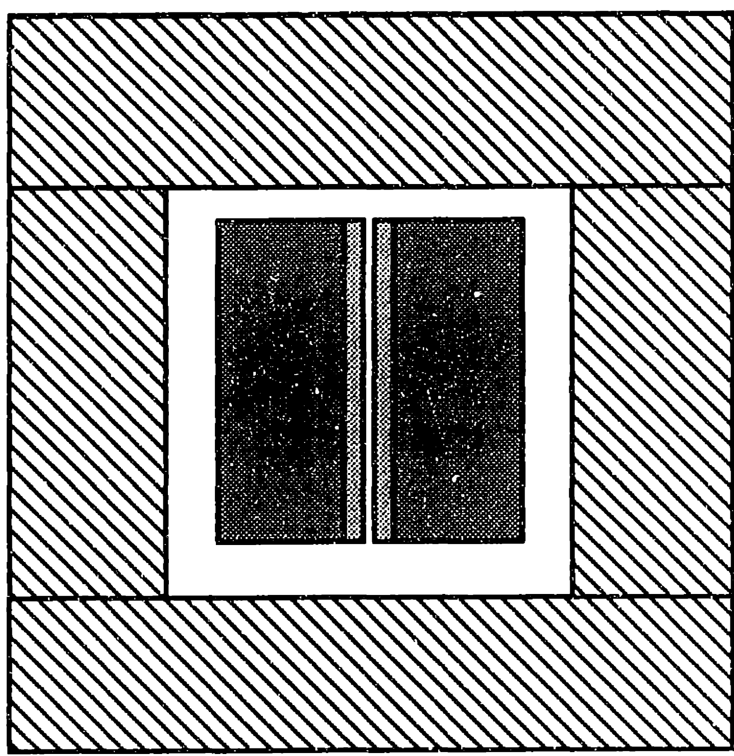
15. Strip all the resist in oxygen plasma.

Profile of the 3 implants are as follows:

<u>Side view:</u>



Top View:



16. Deep boron implant: 1E13 at 180 KeV (For punchthrough control)

17. Channel implant (B or I

18. RCA clean, HF dip.

19. RTA, 1050 C for 5 second in O2 ambiant.

20. RCA Clean, HF dip.

21. Grow the gate oxide.

22. Deposit the Al-Cr-Al gate.

23. -Spin 1 μm positive resist, soft bake.
    -pattern gate.

24. Etch gate in PAN etch, Cr etch and PAD etch.

25. strip the resist in O2 plasma.

26. Deposit 0.6 um of PECVD. of SiO2.

27. Spin 1 μm positive resist, soft bake.

28. Pattern for contact holes, hard bake resist.

29. Descume using oxygen plasma (Etch 0.2 μm of resist).

25. Open contact holes by etching SiO2 in PAD Etch.

30. -Just before sputtering, etch the native oxide in PAD etch.
    -Strip resist in A-20 stripper. Immediately load into sputterer.
    -Sputter 1 um of Al-Si.

31. Spin 1 μm positive resist, soft bake.
    Pattern for contact pads.

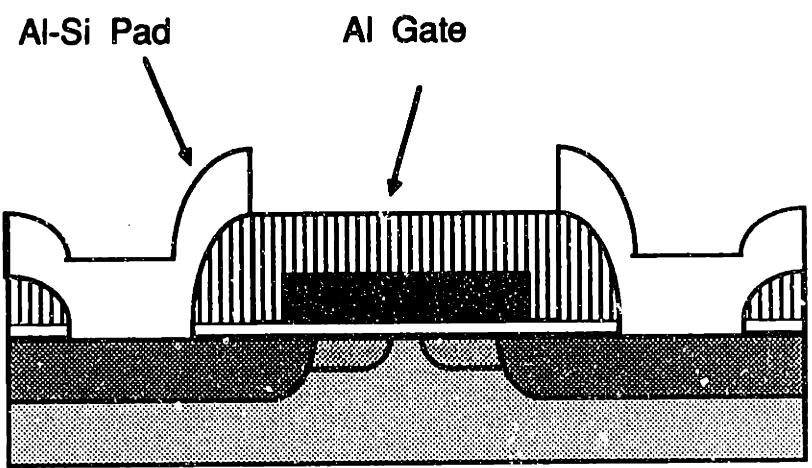
32. Etch in PAN Etch.

33. Strip resist in Oxygen plasma.

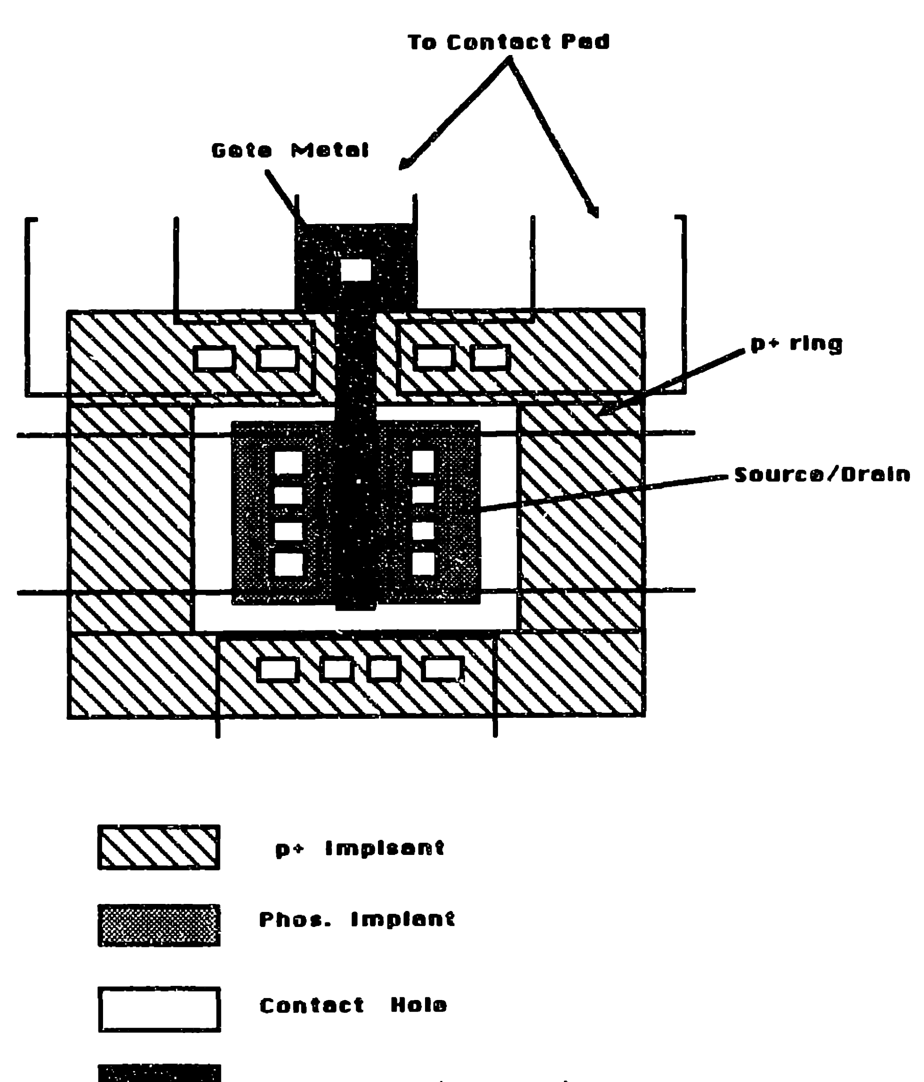
34. Put Al on back.

35. Cut into small chips.

36. Sinter at 400 C in forming gas for 10 min.



SIDE VIEW OF DEVICE



TOP VIEW OF DEVICE

References

- D. A. Antoniadis, I. Moskowitz "Diffusion of Indium in Silicon Inert and Oxidizing Ambient" *J. Appl. Phys.* , Vol 53, 9214, 1982.

- M. Artaki "Hot Electron Flow in an Inhomogeneous Field" *Appl. Phys. Lett.* , Vol 52, 141, 1988.

- G. Baccarani, M. R. Wordeman, "An Investigation of Steady-State Velocity Overshoot in Silicon" *Solid State Elec.* , Vol 28, 407, 1985.

- G. Baccarani, M. R. Wordeman, R. H. Dennard, "Generalized scaling Theory and its Application to 1/4 Micrometer MOSFET Design" *IEEE Trans. Elec. Dev.* , Vol ED-31, 452, 1984.

- G. A. Baraff, "Distribution Functions and Ionization Rates for Hot Electrons in Semiconductors" *Phy. Rev.*, Vol. 128, 2507, 1962.

- K. Blotekjaer, "Transport Equations for Electrons in Two Vally Semiconductors" *IEEE Trans. Elec. Dev.* , Vol ED-17, 38, 1970.

- T. Y. Chan, P. K. Ko, and C. Hu, "A Simple Method to Charactrize Substrate Current in MOSFET's", *IEEE Elec. Dev. Lett.*, EDL-5, 505, 1984.

- S. Y. Chou, D. A. Antoniadis, H. I. Smith, " Observation of Electron Velocity Overshoot in Sub-100-nm-Channel MOSFET's in Si.", *IEEE Elec. Dev. Lett.*, EDL-6, 665, 1985.

- S. Y. Chou, H. I. Smith and D. A. Antoniadis, "Sub-100nm Channel Length Transistors Fabricated Using X-ray Lithography", J. Vac. Sci. Tech. B, 253, 1986.

- S. Y. Chou and D. A. Antoniadis, "Relationship Between Measured and Intrinsic Transconductances of FET's" , *IEEE Trans. on Elec. Dev.*, ED-34, 448, 1987.

- E. M. Conwell, "High Field Transport in Semiconductors", Academic Press, New York, 1967.

- C. R. Crowell and S. M. Sze, "Temperature Dependence of Avalanche Multipication in Semiconductors", *Appl. Phy. Lett.*, Vol. 9, 242, 1966. - Y. A. El-Mansy, A. R. Boothroyd, "A Simple Two-Dimensional Model For IGFET Operation in the Saturation Region", *IEEE Trans. on Elec. Dev.*, ED-24, 254, 1977.

- M. Fukuma, and W. W. Lui, " MOSFET Substrate Current Model Including Energy Transport" *IEEE Elec. Dev. Let.* , Vol EDL-8, 214, 1987.

- H. Gesch, J. P. Leburn, and G. E. Dorda, "Generation of Interface States by Hot Hole Injection in MOSFET's", *IEEE Trans. Elec. Devices*, ED-29, 913, 1982.

- A. K. Henning, N. N. Chan, J. T. Watt, and, J. D. Plummer, "Substrate Currents at cryogenic Temperatures: Measurements and a Two Dimensional Model for CMOS Technology" *IEEE Trans. Elec. Dev.* , Vol ED-34, 64, 1987.

- R. S. Huang, P. H. Ladbrooke, "The Physics of Excess Electron Velocity in Submicron Channel FET's", *J. of Appl. Phy.*, Vol 48, 4791, 1977.

- F. Hsu, P. K. Ko, S. Tam, C, Hu, and R. S. Muller, "An Analytical Breakdown Model for Short Channel MOSFET's", *IEEE Trans. Elec. Devices*, ED-29, 1735, 1982.

- C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-Electron-Induced MOSFET Degredation - Model, Monitor, and Improvement", *IEEE Trans. Elec. Devices*, ED-32, 375, 1985.

- P. K. Ko, R. S. Muller, and C. Hu, "A Unified Model For Hot-Electron Currents in MOSFET's", *IEDM Tech. Digest*, 600, 1980.

- T. Kobayashi and K. Saito, "Two-Dimentional Analysis of Velocity Overshoot Effects in Ultrashort-Channel Si MOSFET's", *IEEE Trans. on Elec. Dev.* , ED-32, 788, 1985.

- D. Lau, G. Gildenblat, C. G. Sodini, and D. E. Nelson, "Low Temperature Substrate Current Characterization of N-Channel MOSFET's", *IEDM Tech. Digest*, 565, 1985.

- M.Liang, J. Y. Choi, P. K. Ko, and C. Hu, "Inversion-Layer Capacitance and Mobility of Very Thin Gate Oxide MOSFET's" *IEEE Trans. Elec. Dev.* , Vol ED-33, 409, 1986.

- D. Long, and J. Myers, "Ionized Impurity Scattering Mobility of Electrons in Silicon" *Phy. Rev.* , Vol 115, 1107, 1959.

- T. H. Ning, P. W. Cook, R. H. Dennard, C. M. Osburn, S. E. Schuster, and H. N. Yu, "1 μm MOSFET VLSI Technology: Part VI - Hot Electron Design Constraints", *IEEE Trans. Elec. Devices*, ED-26, 364, 1979.

- J. P. Nougier, J. C. Vaissiere, and D. Gasquet, "Determination of Transient Regime of Hot Carriers in Semiconductors, Using the Relaxation Time Approximation", *J. of Appl. Phy.*, Vol. 52, 825, 1981.

- P. J. Price, "On the Flow Equation in Device Simulation", *Private Communication.*

- E. Rosencher, "A Displaced Maxwellian Approach to Ballistic Electron Transport in Semiconductors", *Solid State Comm.*, Vol. 38, 1293, 1981.

- J. G. Ruch, "Electron Dynamics in Short Channel Field-Effect Transistors", *IEEE Trans. Electron Devices*, ED-19, 652, 1972.

- S. Selberherr, A. Schutz, and H. W. Potzel, "MINIMOS - A Two Dimensional MOS Transistor Analyzer" , *IEEE Trans. on Elec. Dev.*, ED-27, 1540, 1980.

- G. G. Shahidi, D. A. Antoniadis, and Henry I. Smith, "Observation of Electron Velocity Overshoot at Room Temperature and 77 K in Silicon Inversion Layer", *IEEE Elec. Dev. Lett.*, EDL-9, 94, 1988; also *J. of Vac. Sci. Tech. B* , Vol 6, 137, 1988;

- G. G. Shahidi, D. A. Antoniadis, and Henry I. Smith, "Reduction of Channel-Hot-Electron-Generated Substrate Current in Sub-150 nm Channel Length Si MOSFETs", accepted for publication in *IEEE Elec. Dev. Lett.*.

- W. Shockley, "Problems Related to p-n Junctions", *Solis State Elec.*, 35, 1961.

- C. G. Sodini, T. Ekstedt, J. L. Moll, "Charge Accumulation and MOBILITY in Thin Dielectric MOS Transistors" *Solid State Elec.* , Vol 25, 833, 1982.

- C. G. Sodini, P. K. Ko, and J. L. Moll, " The Effect Of High Fields on MOS Device and Circuit Performance" *IEEE Trans. Elec. Dev.* , Vol ED-31, 1386, 1984.

- R. Stratton, "Diffusion of Hot and Cold Electrons in Semoconductor Barriers" *Phys. Rev..* , Vol 126, 2002, 1962.

- P. I. Suciu and R. L. Johnston, *IEEE Trans. on Elec. Dev.*, ED-27, 1846, 1980.

- S. C. Sun and J. D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces", *IEEE Trans. on Elec. Dev.*, ED-27, 1497, 1980.

- E. Takeda, H. Kume, T. Toyabe, and S. Asai, "Submicron MOSFET Structure for Minimizing Hot-Carrier Generation", *IEEE Trans. Elec. Devices*, ED-29, 611, 1982.

- E. Takeda and N. Suzuki, "An Empirical Model for Device Degredation Due to Hot-Carrier Injection",*IEEE Elec. Dev. Lett.*, EDL-4, 111, 1983.

- S. Tam, P. K. Ko, and C. Hu, "Lucky-Electron Model of Channel Hot Electron Injection in MOSFET's", *IEEE Trans. Elec. Devices*, ED-31, 1116, 1984.

- G. W. Taylor, "Velocity-Saturated Characteristics of Short Channel MOSFET's", *AT& T Bell Lab. Tech. J.*, Vol. 63, 1325, 1984.